EXPRESS MAIL NO. EL852691498US

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 February 2001 (08.02.2001)

PCT

(10) International Publication Number
WO 01/09162 A2

(51) International Patent Classification[7]: C07H 21/04, C07K 14/00, 16/00

(21) International Application Number: PCT/US00/20935

(22) International Filing Date: 31 July 2000 (31.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/365,164 30 July 1999 (30.07.1999) US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors: FRASER, Christopher, C.; 52 Grassland Street, Lexington, MA 02421 (US). SHARP, John, D.; 245 Park Avenue, Arlington, MA 02476 (US). KIRST, Susan, J.; 12 Dana Street, Brookline, MA 02445 (US). BARNES, Thomas, M.; 22 Hanson Street #2, Boston, MA 02118 (US). WRIGHTON, Nick; 18 Lloyd Street, Winchester, MA 01890 (US). MYERS, Paul, S.; 14 Cornelius Way, Cambridge, MA 02141 (US). PAN, Yang; 6 Hamilton Road #1, Brookline, MA 02446 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/09162 A2

(54) Title: SECRETED PROTEINS AND USES THEREOF

(57) Abstract:

SECRETED PROTEINS AND USES THEREOF

Cross Reference to Related Applications

This application is a continuation-in-part of U.S. patent application Serial No. 09/365,164, filed July 30, 1999, the contents of which are incorporated herein by reference in its entirety.

Background of the Invention

Many secreted proteins, for example, cytokines and cytokine receptors, play a vital role in the regulation of cell growth, cell differentiation, and a variety of specific cellular responses. A number of medically useful proteins, including erythropoietin, granulocyte-macrophage colony stimulating factor, human growth hormone, and various interleukins, are secreted proteins. Thus, an important goal in the design and development of new therapies is the identification and characterization of secreted and transmembrane proteins and the genes which encode them.

Many secreted proteins are receptors which bind a ligand and transduce an intracellular signal, leading to a variety of cellular responses. The identification and characterization of such a receptor enables one to identify both the ligands which bind to the receptor and the intracellular molecules and signal transduction pathways associated with the receptor, permitting one to identify or design modulators of receptor activity, *e.g.*, receptor agonists or antagonists and modulators of signal transduction.

Summary of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules which encode the TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 proteins, all of which are either wholly secreted or transmembrane proteins.

The TANGO 339 proteins are transmembrane 4 domain-containing polypeptides that exhibit homology to human CD9 antigen, a cell surface antigen associated with platelet activation and aggregation.

The TANGO 353, TANGO 358, and TANGO 365 proteins are transmembrane proteins.

The TANGO 368 proteins are secreted proteins encoded by sequences with homology to genomic sequences of the human T-cell receptor gamma V1 gene region.

The TANGO 383 proteins are transmembrane polypeptides with homology to retinopathy proteins.

The TANGO 393 protein are transmembrane proteins with homology to proteins containing Leucine-rich repeats (LRR) such as the Leucine-Rich α-2-Glycoprotein (LRG), SLIT-1, and Platelet Glycoprotein V (GPV) precursor.

The TANGO 402 proteins are homologous to the LOX-1 protein, which has been associated with low density lipoprotein metabolism and atherosclerosis.

The MANGO 346, MANGO 349, and TANGO 369 proteins are secreted proteins.

The TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 proteins, fragments, derivatives, and variants thereof are collectively referred to herein as "polypeptides of the invention" or "proteins of the invention." Nucleic acid molecules encoding the polypeptides or proteins of the invention are collectively referred to as "nucleic acids of the invention."

The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present invention also provides nucleic acid molecules which are suitable for use as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, the nucleotide sequence of the cDNA insert of an EpT339 clone deposited with the American Type Culture Collection (ATCC®) as Accession Number PTA-292, or a complement thereof.

The invention features nucleic acid molecules, which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:27, SEQ ID NO:28, the nucleotide sequence of the cDNA insert of an EpT353 clone deposited with the ATCC® as Accession Number PTA-292, or a complement thereof.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:36, SEQ ID NO:37, the nucleotide sequence of the cDNA insert of an EpT358 clone deposited with the ATCC® as Accession Number PTA-292, or a complement thereof.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:44, SEQ ID NO:45, the nucleotide sequence of the cDNA insert of an

EpT365 clone deposited with the ATCC® as Accession Number PTA-291, or a complement thereof.

The invention features nucleic acid molecules, preferably cDNA molecules, which are at least 98% identical to the nucleotide sequence of SEQ ID NO:52, SEQ ID NO:53, the nucleotide sequence of the cDNA insert of an EpT368 clone deposited with the ATCC® as Accession Number PTA-291, or a complement thereof.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:58, SEQ ID NO:59, the nucleotide sequence of the cDNA insert of an EpT369 clone deposited with the ATCC® as Accession Number PTA-295, or a complement thereof.

The invention features nucleic acid molecules which are at least 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:63, the nucleotide sequence of the cDNA insert of an EpT383 clone deposited with the ATCC® as Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules which are at least 65%, 75%, 80%, 85%, 90%, 95% or 98% identical to the nucleotide sequence of SEQ ID NO:64, the nucleotide sequence of the cDNA insert of an EpT383 clone deposited with the ATCC® as Accession Number PTA-295, or a complement thereof.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:73, SEQ ID NO:74, the nucleotide sequence of the cDNA insert of a human EpT393 clone deposited with the ATCC® as Accession Number PTA-295, or a complement thereof.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:93, SEQ ID NO:94, the nucleotide sequence of a mouse EpT393 cDNA, or a complement thereof.

The invention features nucleic acid molecules which are at least 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:110, SEQ ID NO:111, the nucleotide sequence of the cDNA insert of an EpT402 clone deposited with the ATCC® as Accession Number PTA-294, or a complement thereof.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:123, SEQ ID NO:124, the nucleotide sequence of the cDNA insert of an

EpM346 clone deposited with the ATCC® as Accession Number PTA-291, or a complement thereof.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:128, SEQ ID NO:129, the nucleotide sequence of the cDNA insert of an EpM349 clone deposited with the ATCC® as Accession Number PTA-295, or a complement thereof.

The invention features nucleic acid molecules which are at least 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, a complement thereof, or the non-coding strand of TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295, wherein the polypeptides or proteins also exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention features nucleic acid molecules of at least 480, 500, 550, 600, 650, 700, 750, 800, 850, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600 or 2700 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:1, the nucleotide sequence of an EpT339 cDNA of ATCC® Accession Number PTA-292, or a complement thereof. The invention also features nucleic acid molecules comprising at least 20, 50, 100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100 contiguous nucleotides of nucleic acids 1 to 2102 of SEQ ID NO:1, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900 or 1000 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2, or a complement thereof.

The invention features nucleic acid molecules of at least 575, 600, 650, 700, 800, 900, 1000, 1100 or 1200 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:27, the nucleotide sequence of an EpT353 cDNA of ATCC® Accession Number PTA-292, or a complement thereof. The invention also features nucleic acid molecules

comprising at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 or 630 contiguous nucleotides of nucleic acids 1 to 634 of SEQ ID NO:27, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 or 690 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:28, or a complement thereof. The invention also features nucleic acid molecules comprising at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 or 560 contiguous nucleotides of nucleic acids 1 to 560 of SEQ ID NO:28, or a complement thereof.

The invention features nucleic acid molecules of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1100, 1200, 1300, 1400, 1500 or 1600 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:36, the nucleotide sequence of an EpT358 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200 or 240 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:37, or a complement thereof.

The invention features nucleic acid molecules of at least 20, 50, 100, 150, 200, 250, 300, 340, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1100, 1200 or 1300 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:44, the nucleotide sequence of an EpT365 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, or 450 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:45, or a complement thereof.

The invention features nucleic acid molecules, preferably a cDNA molecule, which are at least 970 nucleotides of the nucleotide sequence the of the nucleotide sequence of SEQ ID NO:44, the nucleotide sequence of an EpT368 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

The invention features nucleic acid molecules of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1050 or 1100 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:58, the nucleotide sequence of an EpT369 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150 or 174 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:59, or a complement thereof.

The invention features nucleic acid molecules of at least 510, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1100, 1200 or 1300 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:63, the nucleotide sequence of an EpT383 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 270, 300, 350, or 400 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:64, or a complement thereof.

The invention also features nucleic acid molecules comprising at least 20, 50, 100, 150, 200 or 250 contiguous nucleotides of nucleic acids 1 to 255 of SEQ ID NO:64, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 or 610 contiguous nucleotides of nucleic acids 775 to 1386 of SEQ ID NO:63, or a complement thereof.

The invention features nucleic acid molecules of at least 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600 or 1700 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:73, the nucleotide sequence of a human EpT393 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, or 1400 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:74, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200 or 1250 contiguous nucleotides of nucleotides 1 to 1250 of SEQ ID NO:73, or a complement thereof.

The invention features nucleic acid molecules of at least 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, or 1900 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:93, the nucleotide sequence of a mouse EpT393 cDNA, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 or 984 contiguous nucleotides nucleotides of nucleic acids 1 to 984 of SEQ ID NO:93, or a complement thereof. The invention also features nucleic acid molecules

which include a fragment of at least 20, 50, 100, 150, 200, 250 or 292 contiguous nucleotides of the nucleic acids 1177 to 1469 of SEQ ID NO:93, or a complement thereof. The invention also features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250 or 280 contiguous nucleotides of the nucleic acids 1666 to 1946 of SEQ ID NO:93, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, or 1400 contiguous nucleotide of the nucleotide sequence of SEQ ID NO:94, or a complement thereof.

The invention features nucleic acid molecules of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1100, 1200 or 1300 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:110, the nucleotide sequence of an EpT402 cDNA of ATCC® Accession Number PTA-294, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 or 620 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:111, or a complement thereof.

The invention features nucleic acid molecules of at least 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1050, 1100 or 1150 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:123, the nucleotide sequence of an EpM346 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

The invention features nucleic acid molecules of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 or 800 contiguous nucleotides of nucleic acids 1 to 805 of SEQ ID NO:123, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 50, 75, 100, or 150 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:124, or a complement thereof.

The invention features nucleic acid molecules of at least 20, 50, 100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500 or 3600 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:128, the nucleotide sequence of an EpM349 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450 or 500 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:129, or a complement thereof.

The invention features isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 20, 50, 100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650 or more contiguous nucleotides identical to the nucleic acid sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or a complement thereof, or the non-coding strand of TANGO 339, TANGO 353, TANGO, 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, wherein the polypeptides or proteins also exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:3, the amino acid sequence encoded by an EpT339 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:29, the amino acid sequence encoded by an EpT353 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:38, the amino acid sequence encoded by an EpT358 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 60%, 65%, 70%, 80%, 88%, 90%, 95% or 98% identical to the amino acid sequence of SEQ ID NO:46, the amino acid sequence encoded by an EpT365 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

The invention also features nucleic acid molecules, preferably cDNA molecules, which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 99% identical to the amino acid sequence of SEQ ID NO:54, the amino acid

sequence encoded by an EpT368 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:60, the amino acid sequence encoded by an EpT369 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 65%, 70%, 80%, 88%, 90%, 95% or 98% identical to the amino acid sequence of SEQ ID NO:65, the amino acid sequence encoded by an EpT383 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 60%, 65%, 70%, 80%, 88%, 90%, 95% or 98% identical to the amino acid sequence of SEQ ID NO:75, the amino acid sequence encoded by a human EpT393 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 60%, 65%, 70%, 80%, 88%, 90%, 95% or 98% identical to the amino acid sequence of SEQ ID NO:95, the amino acid sequence encoded by a mouse EpT393 cDNA, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 26%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:112, the amino acid sequence encoded by an EpT402 cDNA of ATCC® Accession Number PTA-294, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 60%, 65%, 70%, 80%, 88%, 90%, 95% or 98% identical to the amino acid sequence of SEQ ID NO:125, the amino acid sequence encoded by an EpM346 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25%, 30%, 35%, 40%, 45%, 50%, 60%, 65%, 70%, 80%, 88%, 90%, 95% or 98% identical to the

amino acid sequence of SEQ ID NO:130, the amino acid sequence encoded by an EpM349 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 25% preferably 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125 or 130, the amino acid sequence encoded by TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, or a complement thereof, wherein the protein encoded by the nucleotide sequence also exhibits at least one structural and/or functional feature of a polypeptide of the invention.

In preferred embodiments, the nucleic acid molecules have the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or the nucleotide sequence of the cDNA clones of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:3, or a fragment thereof including at least 10, 15, 20, 25, 30, 50, 75, 100, 125, 150, 175, 200, 225, 250, 260 or 265 contiguous amino acids of SEQ ID NO:3, or the amino acid sequence encoded by an EpT339 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:29, or a fragment thereof including at least 45, 50, 75, 100, 125, 150, 175, 200 or 225 contiguous amino acids of SEQ ID NO:29, or the amino acid sequence encoded by an EpT353 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:38, or a fragment thereof including at least 10, 15, 20, 25, 30, 50, 75, or 80 contiguous amino acids of SEQ ID NO:38, or the amino acid sequence encoded by an EpT358 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:46, or a fragment thereof including at least 10, 15, 20, 25, 30, 50, 75, 90, 100, 125, 150 or 160 contiguous amino acids of SEQ ID NO:46, or the amino acid sequence encoded by an EpT365 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

Also within the invention are nucleic acid molecules, preferably cDNA molecules, which encode a polypeptide having the amino acid sequence of SEQ ID NO:54, or a fragment thereof including at least 10, 15, 20, 25, 30, 50, 60, 70, 75 or 80 contiguous amino acids of SEQ ID NO:54, or the amino acid sequence encoded by an EpT368 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:60, or a fragment thereof including at least 10, 15, 20, 25, 30, 50 or 55 contiguous amino acids of SEQ ID NO:60, or the amino acid sequence encoded by an EpT369 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:65, or a fragment thereof including at least 90, 100, 110, 125 or 135 contiguous amino acids of SEQ ID NO:65, or the amino acid sequence encoded by an EpT383 cDNA of ATCC® Accession Number PTA-295, or a complement thereof, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:75, or a fragment thereof including at least 60, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400 or 450 contiguous amino acids of SEQ ID NO:75, or the amino acid sequence encoded by a human EpT393 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:95, or a fragment thereof including at least 10, 15, 20, 25, 30, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400 or 450 contiguous amino acids of SEQ ID NO:95, or the amino acid sequence encoded by a mouse EpT393 cDNA, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:112, or a fragment thereof including at least 10, 15, 20, 25, 30, 50, 75, 100, 125, 150, 175 or 200 contiguous amino acids of SEQ ID NO:112, he amino acid sequence encoded by an EpT402 cDNA of ATCC® Accession Number PTA-294, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:125, or a fragment thereof including at least 10, 15, 20, 25, 30, 35, 40, 45, 50 or 55 contiguous amino acids of SEQ ID NO:125, or the amino acid sequence encoded by an EpM346 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a polypeptide having the amino acid sequence of SEQ ID NO:130, or a fragment thereof including at least 10, 15, 20, 25, 30, 50, 75, 100, 125, 150 or 160 contiguous amino acids of SEQ ID NO:130, or the amino acid sequence encoded by an EpM349 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules which encode a polypeptide fragment of at least 10, 15, 25, 30, 50, 75, 100, 125, 150, 175, 200 or more contiguous amino acids of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125 or 130, or the amino acid sequence encoded by TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, or a complement thereof, wherein the fragment exhibits at least one structural and/or functional feature of a polypeptide of the invention.

The invention includes nucleic acid molecules which encode an allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125 or 130, or the amino acid sequence encoded by a cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, or a complement thereof under stringent conditions.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 30%, preferably 40%, 45%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:3, or the amino acid sequence encoded by an EpT339 cDNA of ATCC® Accession Number PTA-292.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 30%, preferably 40%, 45%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:29, or the amino acid sequence encoded by an EpT353 cDNA of ATCC® Accession Number PTA-292.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 30%, preferably 40%, 45%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:38, or the amino acid sequence encoded by an EpT358 cDNA of ATCC® Accession Number PTA-292.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 30%, preferably 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:46 or the amino acid sequence encoded by an EpT365 cDNA of ATCC® Accession Number PTA-291.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least 30%, preferably 40%, 45%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:54, or the amino acid sequence encoded by an EpT368 cDNA of ATCC® Accession Number PTA-291.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least 30%, preferably 40%, 45%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:60, or the amino acid sequence encoded by an EpT369 cDNA of ATCC® Accession Number PTA-295.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 65%, preferably 70%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:65 or the amino acid sequence encoded by an EpT383 cDNA of ATCC® Accession Number PTA-295.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 25%, preferably 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:75 or the amino acid sequence encoded by a human EpT393 cDNA of ATCC® Accession Number PTA-295.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 25%, preferably 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:95 or the amino acid sequence encoded by a mouse EpT393 cDNA.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 26%, preferably 30%, 40%, 45%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:112, or the amino acid sequence encoded by an EpT402 cDNA of ATCC® Accession Number PTA-294.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 30%, preferably 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:125 or the amino acid sequence encoded by an EpM346 cDNA of ATCC® Accession Number PTA-291.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 30%, preferably 40%, 45%, 50%, 55%, 65%, 75%,

85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:130 or the amino acid sequence encoded by an EpM349 cDNA of ATCC® Accession Number PTA-295.

The invention also features isolated polypeptides or proteins having an amino acid sequence that is at least about 25%, preferably 30%, 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95% or 98% identical to the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 52, 54, 60, 65, 75, 95, 112, 125 or 130, or the amino acid sequence encoded by TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, wherein the polypeptides or proteins also exhibit at least one structural and/or functional feature of a polypeptide of the invention.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:3, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or 2, a complement thereof, or the non-coding strand of an EpT339 cDNA of ATCC® Accession Number PTA-292.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:29, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:27 or 28, a complement thereof, or the non-coding strand of an EpT353 cDNA of ATCC® Accession Number PTA-292.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:38, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:36 or 37, a complement thereof, or the non-coding strand of an EpT358 cDNA of ATCC® Accession Number PTA-292.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:46, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:45, a complement thereof, or the non-coding strand of an EpT365 cDNA of ATCC® Accession Number PTA-291.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:54, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:52 or 53, a complement thereof, or the non-coding strand of an EpT368 cDNA of ATCC® Accession Number PTA-291.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:60, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:58 or 59, a complement thereof, or the non-coding strand of an EpT369 cDNA of ATCC® Accession Number PTA-295.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 65%, preferably 70%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:65, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:64, a complement thereof, or the non-coding strand of an EpT383 cDNA of ATCC® Accession Number PTA-295.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:75, and isolated polypeptides or proteins

which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:74, a complement thereof, or the non-coding strand of a human EpT393 cDNA of ATCC® Accession Number PTA 295.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:96, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:95, a complement thereof, or the non-coding strand of a mouse EpT393 cDNA.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about preferably 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:112, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:110 or 111, a complement thereof, or the non-coding strand of an EpT402 cDNA of ATCC® Accession Number PTA-294.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:125, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:124, a complement thereof, or the non-coding strand of an EpM346 cDNA of ATCC® Accession Number PTA-291.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to the nucleic acid sequence encoding SEQ ID NO:130, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:129, a complement thereof, or the non-coding strand of an EpM349 cDNA of ATCC® Accession Number PTA-295.

The invention features isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 30%, preferably 35%, 40%, 45%, 50%, 55%, 65%, 75%, 85%, 95%, or 98% identical to the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, a complement thereof, or the non-coding strand of TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295, wherein the polypeptides or proteins also exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention also features isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 25%, preferably 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95% or 98% identical to a nucleic acid sequence encoding SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, a complement thereof, or the non-coding strand of TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, wherein the polypeptides or proteins also exhibit at least one structural and/or functional feature of a polypeptide of the invention.

Also within the invention are polypeptides which are allelic variants of a polypeptide that includes the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125 or 130, or the amino acid sequence encoded by a cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294

or Accession Number PTA-295, respectively, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule having the sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, or 129, or a complement thereof under stringent conditions.

Also within the invention are polypeptides which are allelic variants of a polypeptide that includes the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125 or 130, or the amino acid sequence encoded by a cDNA of ATCC® Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, respectively, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule having the sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, or 129, or a complement thereof under stringent conditions, wherein the polypeptides or proteins also exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or 2, or an EpT339 cDNA of ATCC® Accession Number PTA-292, or a complement thereof. In one embodiment, the nucleic acid molecules are at least 480, 500, 550, 600, 650, 700, 750, 800, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600 or 2700 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, an EpT339 cDNA of ATCC® Accession Number PTA-292, or a complement thereof. In another embodiment, the nucleic acid molecules are at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900 or 1000 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2 or nucleic acids 1 to 2100 of SEQ ID NO:1, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:27, or an EpT353 cDNA of ATCC® Accession Number PTA-292, or a complement thereof. In one embodiment, the nucleic acid molecules are at least 575, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150 or 1200 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:27 or 28, an EpT353 cDNA of ATCC® Accession Number PTA-292, or a complement thereof. In another embodiment, the nucleic acid molecules are at least 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 or 690 contiguous nucleotides

in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:28, an EpT353 cDNA of ATCC® Accession Number PTA-292, or a complement thereof. In yet another embodiment, the nucleic acid molecules are at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 or 560 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising nucleotides 1 to 500 of SEQ ID NO:28, an EpT353 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:36 or 37, or an EpT358 cDNA of ATCC® Accession Number PTA-292, or a complement thereof. In one embodiment, the nucleic acid molecules are at least 50, 100, 150, 200 or 240 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:36 or 37, an EpT358 cDNA of ATCC® Accession Number PTA-292, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:44, or an EpT365 cDNA of ATCC® Accession Number PTA-291, or a complement thereof. In one embodiment, the nucleic acid molecules are at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 1000, 1100, or 1150 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:44, an EpT365 cDNA of ATCC® Accession Number PTA-291, or a complement thereof. In another embodiment, the nucleic acid molecules are at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450 or 490 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:45, an EpT365 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:58 or 59, or an EpT369 cDNA of ATCC® Accession Number PTA-295, or a complement thereof. In one embodiment, the nucleic acid molecules are at least 20, 50, 100, 150 or 174 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:58 or 59, an EpT369 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:63, or an EpT383 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

In one embodiment, the nucleic acid molecules are at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, or 600 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotides of SEQ ID NO:63, or an EpT 283 cDNA of ATCC® Accession Number PTA-295, or a complement thereof. Preferably, such nucleic acids hybridize under these conditions to at least a portion of nucleotides 1 to 250 and/or 800 to 1386 of SEQ ID NO:63.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:73, or a human EpT393 cDNA of ATCC® Accession Number PTA-295, or a complement thereof. In other embodiments, the nucleic acid molecules are at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350 or 1386 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotides from of SEQ ID NO:73, a human EpT393 cDNA of ATCC® Accession Number PTA-295, or a complement thereof. Preferably, such nucleic acids hybridize under these conditions to at least a portion of nucleotides 1 to 1250 of SEQ ID NO:73.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:93, or a mouse EpT393 cDNA, or a complement thereof. In one embodiment, the nucleic acid molecules are at least 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800 or 1900 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotides of SEQ ID NO:93, a mouse EpT393 cDNA or a complement thereof. Preferably, such nucleic acids hybridize under these conditions to at least a portion of nucleotides 1 to 950 and/or 1200 to 1800 of SEQ ID NO:93.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:110 or 111, or an EpT402 cDNA of ATCC® Accession Number PTA-294, or a complement thereof. In one embodiment, the nucleic acid molecules are at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 or 620 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:110 or 111, an EpT402 cDNA of ATCC® Accession Number PTA-294, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:123, or an EpM346 cDNA of ATCC® Accession Number PTA-291, or a complement thereof.

In one embodiment, the nucleic acid molecules are at least 450, 500, 550, 600, 650, 700, 750, 800, 1000, 1100 or 1150 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:123, an EpM346 cDNA of ATCC® Accession Number PTA-291, or a complement thereof. In another embodiment, the nucleic acid molecules are at least 50, 75, 100, 125, 150 or 175 contiguous nucleotides in length and hybridize under stringent conditions to the nucleotide sequence of SEQ ID NO:124, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:128, or an EpM349 cDNA of ATCC® Accession Number PTA-295, or a complement thereof. In one embodiment, the nucleic acid molecules are at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, or 3600 contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:128, an EpM349 cDNA of ATCC® Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules at least 15, preferably 50, 75, 100, 150, 200, 250, 300, 350, 400, 500, 600 or more contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or a nucleotide sequence of TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, or a complement thereof.

The invention also features nucleic acid molecules at least 15, preferably 50, 75, 100, 150, 200, 250, 300, 350, 400, 500, 600 or more contiguous nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or a nucleotide sequence of TANGO 339, TANGO 353, TANGO 358, TANGO 365,

TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 cDNA of Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294 or Accession Number PTA-295, or a complement thereof, wherein such nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

In preferred embodiments, the isolated nucleic acid molecules encode a cytoplasmic, transmembrane, or extracellular domain of a polypeptide of the invention.

In one embodiment, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a nucleic acid of the invention.

Another aspect of the invention provides vectors, *e.g.*, recombinant expression vectors, comprising a nucleic acid molecule of the invention. In another embodiment, the invention provides host cells containing such a vector or engineered to contain and/or express a nucleic acid molecule of the invention. The invention also provides methods for producing a polypeptide of the invention by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector encoding a polypeptide of the invention such that the polypeptide of the invention is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides of the invention. Preferred proteins and polypeptides possess at least one biological activity possessed by the corresponding naturally-occurring human polypeptide. An activity, a biological activity, or a functional activity of a polypeptide or nucleic acid of the invention refers to an activity exerted by a protein, polypeptide or nucleic acid molecule of the invention on a responsive cell as determined *in vivo* or *in vitro*, according to standard techniques. Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein, or an indirect activity, such as a cellular signaling activity mediated by interaction of the protein with a second protein.

For TANGO 339 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate (this term, as used herein, includes, but is not limited to, stabilize, promote, inhibit or disrupt) the development, differentiation, proliferation and/or activity of immune cells (*e.g.*, B-lymphocyte function); (2) the ability to modulate the development and progression of cancer (*e.g.* lymphomas and/or melanoma-associated cancer); (3) the ability to modulate, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (4) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions; (5) the ability to modulate hematopoietic processes; (6) the ability to modulate platelet activation and aggregation; (7) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (8) the ability to modulate intercellular signaling (*e.g.*,

in the nervous system); (9) the ability modulate the development, differentiation, proliferation and/or activity of neuronal cells and glial cells (*e.g.*, oligodendrocytes and astrocytes); (10) the ability to modulate the development, differentiation and activity of eye structures, such as the retina (*e.g.*, the ability to modulate photoreceptor disk morphogenesis); and (11) the ability to modulate the development of organs, tissues and/or cells in an embryo and/or fetus.

For TANGO 353 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate development, differentiation, proliferation and/or activity of immune cells, such as lymphocytes (*e.g.*, T cells and B cells); (2) ability to modulate cell proliferation, *e.g.*, abnormal cell proliferation;(3) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (4) the ability to modulate intercellular signaling (*e.g.*, in the immune system); (5) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions; and (6) the ability to modulate, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition.

For TANGO 358 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate development, differentiation, maturation, proliferation and/or activity of immune cells such as thymocytes, *e.g.*, T-lymphocytes; (2) the ability to modulate the host immune response; (3) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (4) the ability to modulate intercellular signaling (*e.g.*, in the immune system); (5) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions; and (6) the ability to modulate, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition.

For TANGO 365 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) the ability to modulate the proliferation, differentiation and/or activity of prostate cells; (3) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (4) the ability to modulate intercellular signaling; (5) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions; and (6) the ability to modulate, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition.

For TANGO 368 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate development, differentiation, proliferation and/or activity of cells, such as immune cells, *e.g.*, natural killer cells; (2) the ability to modulate the host immune

response; (3) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (4) the ability to modulate intercellular signaling (*e.g.*, in the immune system); (5) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions; and (6) the ability to modulate, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition.

For TANGO 369 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate development, differentiation, proliferation and/or activity of cells, such as immune cells, *e.g.*, natural killer cells; (2) the ability to modulate the host immune response; (3) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (4) the ability to modulate intercellular signaling (*e.g.*, in the immune system); (5) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions; and (6) the ability to modulate, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition.

For TANGO 383 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) the ability to modulate the proliferation, differentiation and/or activity of prostate cells; (3) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (4) the ability to modulate intercellular signaling; and (5) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions.

For TANGO 393 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) the ability to modulate the proliferation, differentiation and/or activity of hypothalamus cells; (3) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (4) the ability to modulate intercellular signaling; and (5) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions.

For TANGO 402 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate development, differentiation, proliferation and/or activity of immune cells (*e.g.*, leukocytes and macrophages), endothelial cells and smooth muscle cells; (2) the ability to modulate the host immune response; (3) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (4) the ability to modulate the development of organs, tissues and/or cells of the embryo and/or fetus; (5) the ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions; (6) the ability

to modulate atherosclerosis, *e.g.*, the initiation and progression of atherosclerosis; (7) the ability to modulate low-density lipoproteins *e.g.*, the ability to modulate levels, metabolism and/or cellular uptake of oxidized low-density lipoprotein (Ox-LDL), the ability to bind to Ox-LDL, and the ability to modulate Ox-LDL activity in cells; (8) the ability to modulate atherogenesis; and (9) the ability to modulate inflammatory functions *e.g.*, by modulating leukocyte adhesion to extracellular matrix and/or endothelial cells; (10) the ability to bind proteins, *e.g.*, lipoproteins, *e.g.*, low-density lipoproteins, *e.g.*, oxidatively modified low-density lipoproteins; (11) the ability to modulate internalization of proteins, *e.g.*, lipoproteins, *e.g.*, low-density lipoproteins, *e.g.*, oxidatively modified low-density lipoproteins; (12) the ability to modulate degradation, *e.g.*, proteolytic degradation, of proteins, *e.g.*, lipoproteins, *e.g.*, low-density lipoproteins, *e.g.*, oxidatively modified low-density lipoproteins; (13) the ability to modulate, *e.g.*, increase, uptake of proteins, *e.g.*, lipoproteins, *e.g.*, low-density lipoproteins, *e.g.*, oxidatively modified low-density lipoproteins, by cells, *e.g.*, macrophages and muscle cells, *e.g.*, smooth muscle cells; (14) the ability to modulate, *e.g.*, prevent, lipid deposition, *e.g.*, in arteries, and thus modulate, *e.g.*, prevent, intimal thickening; (15) the ability to modulate, *e.g.*, induce or prevent, changes in cells, *e.g.*, transformation of cells (*e.g.*, macrophages and smooth muscle cells) into foam cells and functional alteration of cells (*e.g.*, endothelial cells, *e.g.*, intimal neovascular endothelial cells); (16) the ability to bind and phagocytose cells, *e.g.*, aged and apoptotic cells; (17) the ability to remove debris, *e.g.*, apoptotic cells, from blood vessel walls; (18) the ability to modulate homeostasis, *e.g.*, vascular homeostasis, *e.g.*, by modulating, *e.g.*, preventing the impairment of, nitric oxide production; (19) the ability to modulate, *e.g.*, inhibit, the expression of molecules, *e.g.*, adhesion molecules (*e.g.*, leukocyte adhesion molecules) and growth factors (*e.g.*, smooth-muscle growth factors); (20) the ability to alter, *e.g.*, increase, expression in response to stimuli, *e.g.*, TNF, shear stress, and pathophysiological stimuli relevant to disorders (*e.g.*, atherosclerosis and inflammation); (21) the ability to form, *e.g.*, stabilize, promote, facilitate, inhibit, or disrupt, cell to cell and cell to blood product interaction, *e.g.*, between leukocytes and platelets or leukocytes and vascular endothelial cells; and (22) the ability to recognize large molecules, *e.g.*, carbohydrates.

For MANGO 346 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) ability to modulate cell-cell interactions; (3) the ability to modulate the proliferation, differentiation and/or activity of neural cells; (4) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); (5) the ability to

modulate neural signaling; (6) the ability to modulate intercellular signaling (*e.g.*, in the neural system); and (7) ability to modulate cell-cell interactions and/or cell-extracellular matrix interactions.

For MANGO 349 or modulators thereof, biological activities include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) ability to modulate cell-cell interactions; (3) the ability to modulate the proliferation, differentiation and/or activity of neural cells; (4) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades); and (5) the ability to modulate intercellular signaling (*e.g.*, in the immune system).

In one embodiment, a polypeptide of the invention has an amino acid sequence sufficiently identical to an identified domain of a polypeptide of the invention. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, with a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have or encode a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain or encode a common structural domain having about 60% identity, preferably 65% identity, more preferably 75%, 85%, 95%, 98% identity.

In one embodiment, a TANGO 339 protein includes at least one or more of the following domains: a signal sequence, an extracellular domain, a transmembrane domain, transmembrane 4 domain, a transmembrane 4-like domain, a peripherin/rom-1 domain, a peripherin/rom-1-like domain, and an intracellular or cytoplasmic domain.

In one embodiment, a TANGO 353 protein includes at least one or more of the following domains: a signal sequence, an extracellular domain, a transmembrane domain and an intracellular or cytoplasmic domain.

In one embodiment, an TANGO 358 includes at least one or more of the following domains: a signal sequence, an extracellular domain, a transmembrane domain, and an intracellular or cytoplasmic domain.

In one embodiment, a TANGO 365 protein includes at least one or more of the following domains: a signal sequence, an extracellular domain, at least one transmembrane domain and an intracellular or cytoplasmic domain.

In one embodiment, a TANGO 368 protein includes at least a signal peptide.

In one embodiment, a TANGO 369 protein includes at least a signal peptide.

In one embodiment, a TANGO 383 protein includes at least one or more of the following domains: a signal sequence, at least one transmembrane domain, an intracellular or cytoplasmic domain, and an extracellular domain.

In one embodiment, a TANGO 393 protein includes at least one or more of the following domains: a signal sequence, an extracellular domain, and a transmembrane domain, a leucine-rich repeat domain and an intracellular or cytoplasmic domain.

In one embodiment, a TANGO 402 protein includes at least one or more of the following domains: a signal sequence, an extracellular domain, a C-type lectin domain, a C-type lectin-like domain, a transmembrane domain, and an intracellular or cytoplasmic domain.

In one embodiment, a MANGO 346 protein includes at least a signal sequence.

In one embodiment, a MANGO 349 protein includes at least a signal sequence.

In one embodiment, the isolated polypeptide of the invention lacks both a transmembrane and a cytoplasmic domain. In another embodiment, the polypeptide lacks both a transmembrane domain and a cytoplasmic domain and is soluble under physiological conditions.

The polypeptides of the present invention, or biologically active portions thereof, can be operably linked to a heterologous amino acid sequence to form fusion proteins. The invention further features antibodies that specifically bind a polypeptide of the invention such as monoclonal or polyclonal antibodies.

In addition, the polypeptides of the invention or biologically active portions thereof, or antibodies of the invention, can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides methods for detecting the presence, activity or expression of a polypeptide of the invention in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of the presence, activity or expression such that the presence activity or expression of a polypeptide of the invention is detected in the biological sample.

In another aspect, the invention provides methods for modulating activity of a polypeptide of the invention comprising contacting a cell with an agent that modulates (inhibits or stimulates) the activity or expression of a polypeptide of the invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention. In another embodiment, the agent is a fragment of a polypeptide of the invention or a nucleic acid molecule encoding such a polypeptide fragment.

In another embodiment, the agent modulates expression of a polypeptide of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an mRNA encoding a polypeptide of the invention.

The present invention also provides methods to treat a subject having a disorder characterized by aberrant activity of a polypeptide of the invention or aberrant expression of a nucleic acid of the invention by administering an agent which is a modulator of the activity of a polypeptide of the invention or a modulator of the expression of a nucleic acid of the invention to the subject. In one embodiment, the modulator is a protein of the invention. In other embodiments, the modulator is a polypeptide (*e.g.*, an antibody or a fragment of a polypeptide of the invention), a peptidomimetic, or other small molecule (*e.g.*, a small organic molecule).

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a polypeptide of the invention, (ii) mis-regulation of a gene encoding a polypeptide of the invention, and (iii) aberrant post-translational modification of the invention wherein a wild-type form of the gene encodes a protein having the activity of the polypeptide of the invention.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In general, such methods entail measuring a biological activity of the polypeptide in the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by measuring the expression of the polypeptide or nucleic acid in the presence and absence of the compound.

In another aspect, the invention provides human or non-human antibodies or fragments thereof which specifically bind to a protein of the invention.

In a preferred embodiment, an antibody or a fragment thereof, *i.e.*, human and non-human antibodies or fragments thereof and/or monoclonal antibodies or fragments thereof of the invention, specifically bind to an extracellular domain having the amino acid sequence of SEQ ID NO:20, 21, 32, 41, 51, 89, 109, 112, 115, 136 or 233.

Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be

conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention and instructions for use. Such kits can also comprise an antibody of the invention conjugated to a detectable substance and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

FIGURE 1 depicts the cDNA sequence of human TANGO 339 (SEQ ID NO:1) and the predicted amino acid sequence of human TANGO 339 (SEQ ID NO:3). The open reading frame of SEQ ID NO:1 extends from nucleotide 210 to nucleotide 1019 of SEQ ID NO:1 (SEQ ID NO:2).

FIGURE 2 depicts a hydropathy plot of human TANGO 339. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) and N-glycosylation site (Ngly) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 42 of SEQ ID NO:3; SEQ ID NO:5) on the left from the mature protein (amino acids 43 to 270 of SEQ ID NO:3; SEQ ID NO:4) on the right.

FIGURE 3 depicts an alignment of the amino acid sequence of human CD9 antigen (SEQ ID NO:24; Accession Number NM__001769) and the amino acid sequence of human TANGO 339 (SEQ ID NO:3). The amino acid sequences of human CD9 antigen and human TANGO 339 are 24.1% identical. This alignment was performed using the ALIGN alignment program with a PAM120 scoring matrix, a gap length penalty of 12, and a gap penalty of 4.

FIGURE 4 depicts an alignment of the nucleotide sequence of the coding region of human CD9 antigen (SEQ ID NO:25; Accession Number NM__001769) and the nucleotide sequence of the coding region of human TANGO 339 (SEQ ID NO:1). The nucleotide sequences of the coding regions of human CD9 antigen and human TANGO 339 are 45.9% identical. This alignment was performed using the ALIGN alignment

program with a PAM120 scoring matrix, a gap length penalty of 12, and a gap penalty of 4.

FIGURE 5 depicts the cDNA sequence of human TANGO 353 (SEQ ID NO:27) and the predicted amino acid sequence of human TANGO 353 (SEQ ID NO:29). The open reading frame of human TANGO 353 extends from nucleotide 76 to nucleotide 765 of SEQ ID NO:27 (SEQ ID NO:28).

FIGURE 6 depicts a hydropathy plot of human TANGO 353. Relatively hydrophobic regions of the protein are shown above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 14 of SEQ ID NO:29; SEQ ID NO:31) on the left from the mature protein (amino acids 15 to 230 of SEQ ID NO:29; SEQ ID NO:30) on the right.

FIGURE 7 depicts a cDNA sequence of human TANGO 358 (SEQ ID NO:36) and the predicted amino acid sequence of human TANGO 358 (SEQ ID NO:38). The open reading frame of human TANGO 358 extends from nucleotide 184 to 429 of SEQ ID NO:36 (SEQ ID NO:37).

FIGURE 8 depicts a hydropathy plot of human TANGO 358. Relatively hydrophobic regions of the protein are shown above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 42 of SEQ ID NO:36; SEQ ID NO:40) on the left from the mature protein (amino acids 43 to 82 of SEQ ID NO:36; SEQ ID NO:39) on the right.

FIGURE 9 depicts the cDNA sequence of human TANGO 365 (SEQ ID NO:44) and the predicted amino acid sequence of human TANGO 365 (SEQ ID NO:46). The open reading frame of SEQ ID NO:44 extends from nucleotide 56 to nucleotide 550 (SEQ ID NO:45).

FIGURE 10 depicts a hydropathy plot of human TANGO 365. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (Cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 36 of SEQ ID NO:46; SEQ ID NO:47) preceding the mature protein (amino acids 37 to 165 of SEQ ID NO:46; SEQ ID NO:48) on the right.

FIGURE 11 depicts the cDNA sequence of human TANGO 368 (SEQ ID NO:52) and the predicted amino acid sequence of TANGO 368 (SEQ ID NO:54). The open reading frame of human TANGO 368 extends from nucleotide 152 to nucleotide 328 of SEQ ID NO:52 (SEQ ID NO:53).

FIGURE 12 depicts a hydropathy plot of human TANGO 368. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (Cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 27 of SEQ ID NO:54; SEQ ID NO:56) on the left from the mature protein (amino acids 28 to 59 of SEQ ID NO:54; SEQ ID NO:55) on the right.

FIGURE 13 depicts a local alignment of the nucleotide sequence of full-length human TANGO 368 (SEQ ID NO:52) and a fragment of the human T-cell receptor gamma V1 gene region (Accession Number AF057177; SEQ ID NO:57). The nucleotide sequence of human TANGO 368 and the human T-cell receptor gamma V1 gene region are 99.3 % identical for a 973 bp overlap. This alignment was performed using the LALIGN program with a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4.

FIGURE 14 depicts a cDNA sequence of human TANGO 369 (SEQ ID NO:58) and the predicted amino acid sequence of human TANGO 369 (SEQ ID NO:60). The open reading frame of human TANGO 369 extends from nucleotide 162 to 335 of SEQ ID NO:58 (SEQ ID NO:61).

FIGURE 15 depicts a hydropathy plot of human TANGO 369. Relatively hydrophobic regions of the protein are shown above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 26 of SEQ ID NO:60; SEQ ID NO:62) on the left from the mature protein (amino acids 27 to 58 of SEQ ID NO:60; SEQ ID NO:61) on the right.

FIGURE 16 depicts the cDNA sequence of human TANGO 383 (SEQ ID NO:63) and the predicted amino acid sequence of human TANGO 383 (SEQ ID NO:65). The open reading frame of SEQ ID NO:63 extends from nucleotide 104 to nucleotide 523 (SEQ ID NO:64).

FIGURE 17 depicts a hydropathy plot of human TANGO 383. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine

residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 20 of SEQ ID NO:65; SEQ ID NO:66) on the left from the mature protein (amino acids 21 to 140 of SEQ ID NO:65; SEQ ID NO:67) on the right.

FIGURE 18 depicts an alignment of the amino acid sequence of TANGO 383 (SEQ ID NO:65) and the amino acid sequence of Neuronal Thread Protein AD7C-NTP (SEQ ID NO:72). The alignments demonstrates that the amino acid sequences of TANGO 383 and Neuronal Thread Protein AD7C-NTP are 52% identical. This alignment was performed using the ProDom NCBI-BLASTP2 program with graphical output using the following settings: Matrix: BLOSUM62; Expect: 0.1; Filter: none.

FIGURE 19 depicts the cDNA sequence of human TANGO 393 (SEQ ID NO:73) and the predicted amino acid sequence of human TANGO 393 (SEQ ID NO:75). The open reading frame of SEQ ID NO:73 extends from nucleotide 40 to nucleotide 1458 (SEQ ID NO:74).

FIGURE 20 depicts a hydropathy plot of human TANGO 393. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) and N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 26 of SEQ ID NO:75; SEQ ID NO:76) on the left from the mature protein (amino acids 27 to 473 of SEQ ID NO:75; SEQ ID NO:77) on the right.

FIGURE 21 depicts the cDNA sequence of mouse TANGO 393 (SEQ ID NO:93) and the predicted amino acid sequence of mouse TANGO 393 (SEQ ID NO:95). The open reading frame of SEQ ID NO:93 extends from nucleotide 226 to nucleotide 1644 (SEQ ID NO:94).

FIGURE 22 depicts a hydropathy plot of mouse TANGO 393. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) and N-glycosylation sites (N-Gly) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 26 of SEQ ID NO:95; SEQ ID NO:96) on the left from the mature protein (amino acids 27 to 473 of SEQ ID NO:95; SEQ ID NO:97) on the right.

FIGURE 23 depicts an alignment of the open reading frames of human TANGO 393 (SEQ ID NO:74) and mouse TANGO 393 (SEQ ID NO:94) demonstrating an identity of 82.8%. The algorithm used to align the sequences was the ALIGN program which calculates a global alignment of two sequences. (Version 2.0u, Myers and Miller, 1989)

FIGURE 24 depicts an alignment of the immature proteins of human TANGO 393 (SEQ ID NO:75) and mouse TANGO 393 (SEQ ID NO:95) demonstrating an identity of 89.2%. The algorithm used to align the sequences was the ALIGN program which calculates a global alignment of two sequences. (Version 2.0u, Myers and Miller, 1989)

FIGURE 25 depicts the cDNA sequence of human TANGO 402 (SEQ ID NO:110) and the predicted amino acid sequence of human TANGO 402 (SEQ ID NO:112). The open reading frame of human TANGO 402 extends from nucleotide 87 to nucleotide 707 of SEQ ID NO:110 (SEQ ID NO:111).

FIGURE 26 depicts a hydropathy plot of human TANGO 402. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) and N-glycosylation (Ngly) sites are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 50 of SEQ ID NO:112; SEQ ID NO:114) on the left from the mature protein (amino acids 51 to 207 of SEQ ID NO:112; SEQ ID NO:113) on the right.

FIGURE 27 depicts an alignment of the amino acid sequence of human TANGO 402 (SEQ ID NO:112) and the amino acid sequence of human LOX-1 (SEQ ID NO:122; Accession Number AB010710). The alignment demonstrates that the amino acid sequences of human TANGO 402 and human LOX-1 are 25.1% identical. This alignment was performed using the ALIGN program with a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4.

FIGURE 28 depicts an alignment of the nucleotide sequences of the open reading frames of human TANGO 402 (SEQ ID NO:111) and human LOX-1 (SEQ ID NO:121; Accession Number AB010710). The alignment of the open reading frame of human TANGO 402 and that of human LOX-1 demonstrates that those two coding regions are 42.0 % identical. An alignment demonstrates that the nucleotide sequences of the cDNA of human TANGO 402 and human LOX-1 are 44.0 % identical. The alignments were performed using the ALIGN program with a PAM120 scoring matrix, a gap length penalty of 12, and a gap penalty of 4.

FIGURE 29 depicts the cDNA sequence of human MANGO 346 (SEQ ID NO:123) and the predicted amino acid sequence of human MANGO 346 (SEQ ID NO:125). The open reading frame of SEQ ID NO:123 extends from nucleotide 319 to nucleotide 498 (SEQ ID NO:124).

FIGURE 30 depicts a hydropathy plot of human MANGO 346. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine

residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 19 of SEQ ID NO:125; SEQ ID NO:126) on the left from the mature protein (amino acids 20 to 60 of SEQ ID NO:125; SEQ ID NO:127) on the right.

FIGURE 31 depicts the cDNA sequence of human MANGO 349 (SEQ ID NO:128) and the predicted amino acid sequence of human MANGO 349 (SEQ ID NO:130). The open reading frame of SEQ ID NO:128 extends from nucleotide 221 to nucleotide 721 (SEQ ID NO:129).

FIGURE 32 depicts a hydropathy plot of human MANGO 349. Relatively hydrophobic regions of the protein are shown above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 26 of SEQ ID NO:130; SEQ ID NO:131) on the left from the mature protein (amino acids 27 to 167 of SEQ ID NO:130; SEQ ID NO:132) on the right.

Detailed Description of the Invention

The TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 and MANGO 349 proteins and nucleic acid molecules comprise families of molecules having certain conserved structural and functional features. As used herein, the terms "family" or "families" are intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species. For example, a family can comprises two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin. Members of the same family may also have common structural domains.

For example, TANGO 339 proteins, TANGO 353 proteins, TANGO 358 proteins, TANGO 365 proteins, TANGO 368 proteins, TANGO 369 proteins, TANGO 383 proteins, TANGO 393 proteins, TANGO 402 proteins, MANGO 346 proteins and MANGO 349 proteins of the invention can have signal sequences. As used herein, a "signal sequence" includes a peptide of at least about 15 or 20 amino acid residues in length which occurs at the N-terminus of secretory and membrane-bound proteins and which contains at least about 70% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 40 amino acid residues,

preferably about 19-34 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 339 protein contains a signal sequence of about amino acids 1 to 42 of SEQ ID NO:3 (SEQ ID NO:5).

In another embodiment, a TANGO 353 protein contains a signal sequence of about amino acids 1 to 14 of SEQ ID NO:37 (SEQ ID NO:31). In another embodiment, a TANGO 358 protein contains a signal sequence at about amino acids 1 to 42 of SEQ ID NO:38 (SEQ ID NO:40). In another embodiment, a TANGO 365 protein contains a signal sequence of about amino acids 1 to 36 of SEQ ID NO:46 (SEQ ID NO:47). In another embodiment, a TANGO 368 protein contains a signal sequence of about amino acids 1 to 27 of SEQ ID NO:54 (SEQ ID NO:56). In another embodiment, a TANGO 369 protein contains a signal sequence of about amino acids 1 to 26 of SEQ ID NO:60 (SEQ ID NO:62). In another embodiment, a TANGO 383 protein contains a signal sequence of about amino acids 1 to 20 of SEQ ID NO:65 (SEQ ID NO:66). In another embodiment, human TANGO 393 protein contains a signal sequence of about amino acids 1 to 26 of SEQ ID NO:75 (SEQ ID NO:76). In another embodiment, mouse TANGO 393 protein contains a signal sequence of about amino acids 1 to 26 of SEQ ID NO:95 (SEQ ID NO:96). In another embodiment, a TANGO 402 protein contains a signal sequence of about amino acids 1 to 50 of SEQ ID NO:112 (SEQ ID NO:114). In another embodiment, a MANGO 346 protein contains a signal sequence of about amino acids 1 to 19 of SEQ ID NO:125 (SEQ ID NO:126). In another embodiment, a MANGO 349 protein contains a signal sequence of about amino acids 1 to 26 of SEQ ID NO:130 (SEQ ID NO:131). The signal sequence is usually cleaved during processing of the mature protein. In the case of, *e.g.*, transmembrane 4-type proteins, the signal peptide is generally not cleaved, but becomes a transmembrane-anchoring domain of the polypeptide.

A TANGO 339 family member can include one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain. In one embodiment, a TANGO 339 protein contains extracellular domains at about amino acid residues 43 to 61 and 116 to 232 of SEQ ID NO:3 (SEQ ID NO:20 and SEQ ID NO:21, respectively), transmembrane domains at about amino acid residues 62 to 84, 93 to 115, and 233 to 254 of SEQ ID NO:3 (SEQ ID NO:15, SEQ ID NO:16 and SEQ ID NO:17, respectively), and cytoplasmic domains at about amino acid residues 85 to 92 and 255 to 270 of SEQ ID NO:3 (SEQ ID NO:22 and SEQ ID NO:23, respectively). In this embodiment, the mature TANGO 339 protein corresponds to amino acids 43 to 270 of SEQ ID NO:3 (SEQ ID NO:4).

In another embodiment, a TANGO 339 protein contains extracellular domains at about amino acid residues 1 to 16, 85 to 92, and 255 to 270 of SEQ ID NO:3 (SEQ ID NO:11, SEQ ID NO:12, and SEQ ID NO:13, respectively), transmembrane domains at about amino acid residues 17 to 41, 62 to 84, 93 to 115, and 233 to 254 of SEQ ID NO:3 (SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16 and SEQ ID NO:17, respectively), and cytoplasmic domains at about amino acid residues 42 to 61 and 116 to 232 of SEQ ID NO:3 (SEQ ID NO:18 and SEQ ID NO:19, respectively). In this embodiment, the mature TANGO 339 protein corresponds to amino acids 1 to 270 of SEQ ID NO:3.

A TANGO 339 family member can include a signal sequence. In certain embodiment, a TANGO 339 family member has the amino acid sequence of SEQ ID NO:3, and the signal sequence is located at amino acids 1 to 40, 1 to 41, 1 to 42, 1 to 43 or 1 to 44. In such embodiments of the invention, the domains and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 40 results in an extracellular domain consisting of amino acids 41 to 61 of SEQ ID NO:3 and the mature TANGO 339 protein corresponding to amino 41 to 270.

A TANGO 339 family member can include one or more transmembrane 4 or transmembrane 4-like domains. A transmembrane 4 domain typically has the following consensus sequence: G-xxx-[LIVMF]-xx-[GSA]-[LIVMF][LIVMF]-G-C-x-[GA]-[STA]-xx-[EG]-xx-[CWN]-[LIVM][LIVM], wherein G is a glycine residue, "x" is any amino acid, [LIVMF] is a leucine, isoleucine, valine, methionine or phenylalanine residue, [GA] is either a glycine or an alanine residue, [STA] is a serine, threonine or alanine residue, [EG] is either a glutamic acid or glycine residue, [CWN] is cysteine, tryptophan or asparagine residue. A transmembrane 4 domain is a characteristic of transmembrane 4 superfamily members which include, for example, CD9 antigen, CD37, CD53, CD63, CD81, and CD82. Transmembrane 4 proteins have the following characteristics: they are type III membrane proteins, which contain an N-terminal membrane-anchoring domain that is not cleaved during biosynthesis and that functions both as a translocation signal and as a membrane anchor; they contain a total of four transmembrane domains and at least seven conserved cysteine residues; and they are approximately 218 to 284 amino acid residues.

A transmembrane 4-like domain as described herein can have the following consensus sequence: G-xxx-[LIVMF]-xx-[GSA]-[LIVMF]-x-G-C-x-[GA]-[STA]-xx-[EG]-xx-[CWN]-[LIVM][LIVM], wherein G is a glycine residue, "x" is any amino acid, [LIVMF] is a leucine, isoleucine, valine, methionine or phenylalanine residue, [GA] is either a glycine or an alanine residue, [STA] is a serine, threonine or alanine residue, [EG]

is either a glutamic acid or glycine residue, [CWN] is cysteine, tryptophan or asparagine residue.

In one embodiment, a TANGO 339 family member has the amino acid sequence of SEQ ID NO:3 and, preferably, a transmembrane 4 domain-like consensus sequence is located at about amino acid positions 69 to 91 of SEQ ID NO:3 (SEQ ID NO:7). In another embodiment, a TANGO 339 family member has the amino acid sequence of SEQ ID NO:3 and, preferably, a transmembrane 4-like domain is located at about amino acid positions 68 to 260 of SEQ ID NO:3 (SEQ ID NO:6). In another embodiment, a TANGO 339 family member includes one or more transmembrane 4-like domain consensus sequences having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 69 to 91 of SEQ ID NO:3 (SEQ ID NO:7). In yet another embodiment, a TANGO 339 family member includes one or more transmembrane 4-like domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 68 to 261 of SEQ ID NO:3 (SEQ ID NO:6).

In another embodiment, a TANGO 339 family member includes one or more transmembrane 4-like domain consensus sequences having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 69 to 91 of SEQ ID NO:3 (SEQ ID NO:7), and has at least one TANGO 339 biological activity as described herein. In yet another embodiment a TANGO 339 family member includes one or more transmembrane 4-like domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 68 to 261 of SEQ ID NO:3 (SEQ ID NO:6), and has at least one TANGO 339 biological activity as described herein.

In another embodiment, the transmembrane 4-like domain of TANGO 339 is a transmembrane 4 domain, which has the following consensus sequence: G-xxx-[LIVMF]-xx-[GSA]-[LIVMF][LIVMF]-G-C-x-[GA]-[STA]-xx-[EG]-xx-[CWN]-[LIVM][LIVM], wherein G is a glycine residue, "x" is any amino acid, [LIVMF] is a leucine, isoleucine, valine, methionine or phenylalanine residue, [GA] is either a glycine or an alanine residue, [STA] is a serine, threonine or alanine residue, [EG] is either a glutamic acid or glycine residue, [CWN] is cysteine, tryptophan or asparagine residue. In this embodiment, a TANGO 339 family member includes one or more transmembrane 4-like domains having

an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 68 to 261 of SEQ ID NO:3 (SEQ ID NO:6).

In another embodiment, a TANGO 339 family member includes one or more peripherin/rom-1 or peripherin/rom-1-like domains. A peripherin/rom-1 domain typically has the following consensus sequence: D-G-V-P-F-S-C-C-N-P-x-S-P-R-P-C, wherein D is an aspartic acid residue, G is a glycine residue, V is a valine residue, P is a proline residue, F is a phenylalanine residue, S is a serine residue, C is a cysteine residue, N is an asparagine residue, x is any amino acid, and R is an arginine residue. Peripherin/rom-1 domains are characteristic of retinal-specific integral membrane proteins that are located at the rims of the photoreceptor disks and that function in disk morphogenesis. Peripherin (or RDS) and rom-1 are examples of proteins that contain the peripherin/rom-1 domain. Defects in the peripherin gene have been shown to cause various diseases, including autosomal dominant retinitis pigmentosa, autosomal dominant punctata albescens, and butterfly-shaped pigment dystrophy.

A peripherin/rom-1-like domain as described herein has the following consensus sequence: G-V-P-F-S-C-C-x-P, wherein G is a glycine residue, V is a valine residue, P is a proline residue, F is a phenylalanine residue, and C is a cysteine residue. In one embodiment, a TANGO 339 family member has the amino acid sequence of SEQ ID NO:3 and, preferably, a peripherin/rom-1-like domain consensus sequence is located at about amino acid positions 181 to 189 of SEQ ID NO:3 (SEQ ID NO:9). In another embodiment, a TANGO 339 family member has the amino acid sequence of SEQ ID NO:31 and, preferably, a peripherin/rom-1-like domain is located at about amino acid positions 18 to 270 of SEQ ID NO:3 (SEQ ID NO:8).

In another embodiment, a TANGO 339 family member includes one or more peripherin/rom-1-like domain consensus sequences having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 181 to 189 of SEQ ID NO:3 (SEQ ID NO:9). In another embodiment, a TANGO 339 family member includes one or more peripherin/rom-1-like domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acid positions 18 to 270 of SEQ ID NO:3 (SEQ ID NO:8).

In another embodiment, a TANGO 339 family member includes one or more peripherin/rom-1-like domain consensus sequences having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more

preferably at least about 85%, and most preferably at least about 95% identical to amino acids 181 to 189 of SEQ ID NO:3 (SEQ ID NO:9), and has at least one TANGO 339 biological activity as described herein. In yet another embodiment, a TANGO 339 family member includes one or more peripherin/rom-1-like domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acid positions 18 to 270 of SEQ ID NO:3 (SEQ ID NO:8), and has at least one TANGO 339 biological activity as described herein.

In another embodiment, the peripherin/rom-1-like domain of TANGO 339 is a peripherin/rom-1 domain, which has the following consensus sequence: D-G-V-P-F-S-C-C-N-P-x-S-P-R-P-C, wherein D is an aspartic acid residue, G is a glycine residue, V is a valine residue, P is a proline residue, F is a phenylalanine residue, S is a serine residue, C is a cysteine residue, N is an asparagine residue, x is any amino acid, and R is an arginine residue. In this embodiment, a TANGO 339 family member includes one or more peripherin/rom-1-like domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 18 to 270 of SEQ ID NO:3 (SEQ ID NO:8).

A TANGO 353 family member can include one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain. Thus, in one embodiment, an TANGO 353 protein contains an extracellular domain of about amino acids 1 to 116 of SEQ ID NO:29, or a mature extracellular domain of about amino acids 15 to 116 of SEQ ID NO:29 (SEQ ID NO:32). In another embodiment, a TANGO 353 protein contains a transmembrane domain of about amino acids 117 to 141 of SEQ ID NO:29 (SEQ ID NO:33). In another embodiment, a TANGO 353 protein contains a cytoplasmic domain of about amino acids 142 to 230 of SEQ ID NO:29 (SEQ ID NO:34). In yet another embodiment, a TANGO 353 protein is a mature protein containing an extracellular, transmembrane and cytoplasmic domain of about amino acids 15 to 230 of SEQ ID NO:29 (SEQ ID NO:30).

A TANGO 353 family member can include a signal sequence. In certain embodiments, a TANGO 353 family member has the amino acid sequence of SEQ ID NO:29, and the signal sequence is located at amino acids 1 to 12, 1 to 13, 1 to 14, 1 to 15 or 1 to 16. In such embodiments of the invention, the extracellular domain and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 12 results in an

extracellular domain consisting of amino acids 13 to 116 of SEQ ID NO:29 and the mature TANGO 353 protein corresponding to amino 13 to 230.

A TANGO 358 family member can include one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain. In one embodiment, a TANGO 358 protein contains an extracellular domain at amino acids 1 to about 49 of SEQ ID NO:38 or a mature extracellular domain at about amino acid residues 43 to 49 of SEQ ID NO:38 (SEQ ID NO:41), a transmembrane domain at about amino acid residues 50 to 66 of SEQ ID NO:38 (SEQ ID NO:42), and a cytoplasmic domain at about amino acid residues 67 to 82 of SEQ ID NO:38 (SEQ ID NO:43).

A TANGO 358 family member can include a signal sequence. In certain embodiment, a TANGO 358 family member has the amino acid sequence of SEQ ID NO:38, and the signal sequence is located at amino acids 1 to 40, 1 to 41, 1 to 42, 1 to 43 or 1 to 44. In such embodiments of the invention, the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 40 results in an extracellular domain consisting of amino acids 41 to 50 of SEQ ID NO:38 and the mature TANGO 368 protein corresponding to amino 41 to 82.

A TANGO 365 family member can include a signal sequence. In certain embodiments, a TANGO 365 family member has the amino acid sequence of SEQ ID NO:46, and the signal sequence is located at amino acids 1 to 34, 1 to 35, 1 to 36, 1 to 37 or 1 to 38. In such embodiments of the invention, the extracellular domain and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 36 results in a mature TANGO 365 protein corresponding to amino 37 to 165 of SEQ ID NO:46 (SEQ ID NO:47).

A TANGO 365 family member can include one or more of the following domains: (1) an extracellular domain; (2) two transmembrane domains; and (3) a cytoplasmic domain. Thus, in one embodiment, a TANGO 365 protein contains an extracellular domain of about amino acids 95 to 165 of SEQ ID NO:46 (SEQ ID NO:51), or a mature extracellular domain of about amino acids 30 to 246 of SEQ ID NO:46. In another embodiment, a TANGO 365 protein contains a first transmembrane domain of about amino acids 52 to 70 of SEQ ID NO:46 (SEQ ID NO:49). In another embodiment, an protein contains a cytoplasmic domain of about amino acids 71 to 77 of SEQ ID NO:46 (SEQ ID NO:133). In another embodiment, a TANGO 365 protein contains a second transmembrane domain of about amino acids 78 to 94 of SEQ ID NO:46 (SEQ ID NO:50). In yet another embodiment, a TANGO 365 protein is a mature protein containing an

extracellular domain, two transmembrane domains and a cytoplasmic domain of about amino acids 37 to 165 of SEQ ID NO:46 (SEQ ID NO:48).

A TANGO 368 family member can include a signal sequence. In certain embodiments, a TANGO 368 family member has the amino acid sequence of SEQ ID NO:54, and the signal sequence is located at amino acids 1 to 25, 1 to 26, 1 to 27, 1 to 28 or 1 to 29. In such embodiments of the invention, the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 27 results in a mature TANGO 368 protein corresponding to amino 28 to 59 of SEQ ID NO:54 (SEQ ID NO:55).

A TANGO 369 family member can include a signal sequence. In certain embodiments, a TANGO 369 family member has the amino acid sequence of SEQ ID NO:60, and the signal sequence is located at amino acids 1 to 24, 1 to 25, 1 to 26, 1 to 27 or 1 to 28. In such embodiments of the invention, the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 26 results in a mature TANGO 368 protein corresponding to amino 27 to 58 of SEQ ID NO:60 (SEQ ID NO:61).

A TANGO 383 family member can include a signal sequence. In certain embodiments, a TANGO 383 family member has the amino acid sequence of SEQ ID NO:65, and the signal sequence is located at amino acids 1 to 18, 1 to 19, 1 to 20, or 1 to 21. In such embodiments of the invention, the extracellular domain and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 20 results in a mature TANGO 383 protein corresponding to amino 21 to 140 of SEQ ID NO:65.

A TANGO 383 family member can include one or more of the following domains: (1) an extracellular domain; (2) two transmembrane domains; and (3) a cytoplasmic domain. In one embodiment, a TANGO 383 protein contains a cytoplasmic domain of about amino acids 21 to 49 of SEQ ID NO:65. In another embodiment, a TANGO 383 protein contains a first transmembrane domain of about amino acids 50 to 70 of SEQ ID NO:65 (SEQ ID NO:68). In another embodiment, a TANGO 383 protein contains an extracellular domain of about amino acids 71 to 115 of SEQ ID NO:65 (SEQ ID NO:70). In another embodiment, a TANGO 383 protein contains a second transmembrane domain of about amino acids 116 to 133 of SEQ ID NO:65 (SEQ ID NO:69). In yet another embodiment, a TANGO 383 protein is a mature protein containing an extracellular domain, two transmembrane domains and a cytoplasmic domain of about amino acids 21 to 140 of SEQ ID NO:65 (SEQ ID NO:67).

In another example, a TANGO 393 family member can include one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain; and (4) a leucine-rich domain. In one embodiment, a TANGO 393 protein contains an extracellular domain at amino acids 1 to about 447 of SEQ ID NO:75 or a mature extracellular domain at about amino acid residues 27 to 447 of SEQ ID NO:75 (SEQ ID NO:89), a transmembrane domain at about amino acid residues 448 to 467 of SEQ ID NO:75 (SEQ ID NO:78), and a cytoplasmic domain at about amino acid residues 468 to 473 of SEQ ID NO:75 (SEQ ID NO:134). In another embodiment, a TANGO 393 family member contains an extracellular domain at amino acids 1 to about 26 of SEQ ID NO:95 or a mature extracellular domain at about amino acid residues 27 to 449 of SEQ ID NO:95 (SEQ ID NO:109), a transmembrane domain at about amino acid residues 450 to 467 of SEQ ID NO:95 (SEQ ID NO:98), and a cytoplasmic domain at about amino acid residues 468 to 473 of SEQ ID NO:95 (SEQ ID NO:35).

A TANGO 393 family member can include one or more leucine-rich-repeat (LRR) domains. A leucine-rich-repeat domain typically has the following degenerate consensus sequence: x-L-x-x-L-x-L-x-x-[NCT]-x-L-x-x-x-L-x-x-x-x-L-x-x-L, wherein L is a leucine residue and can be replaced by any aliphatic residue,"x" is any amino acid, and [NCT] is either an asparagine, cysteine or threonine, respectively. Leucine-rich-repeat domains most frequently appear in tandem repeats. The degenerate leucine-rich-repeat domains are characteristic of a diverse set of signaling proteins that are involved in cell signaling, cell growth and cell differentiation. Defects in leucine-rich-repeat genes have been shown to cause various diseases which include but are not limited to Bernard-Soulier disease, a bleeding disorder. Furthermore, leucine-rich-repeat genes are involved in the pathogenesis of diseases, for example, the leucine-rich-repeat of type-1 human immunodeficiency virus Rev protein is the trans-activating region of the virus (Kobe and Deisenhofer, 1994, TIBS, 19:415-421).

In one embodiment, a TANGO 393 family member has the amino acid sequence of SEQ ID NO:76 and, preferably, a leucine-rich-repeat domain consensus sequence is located at about amino acid positions 26 to 57, 58 to 81, 82 to 105, 106 to 130, 131 to 154, 155 to 178, 179 to 202, 203 to 226, 227 to 250, and/or 260 to 310 of human TANGO 393 (SEQ ID NO:76), SEQ ID NO:79, 80, 81, 83, 83, 84, 85, 86, 87 and 88, respectively. In another embodiment, a TANGO 393 family member has the amino acid sequence of SEQ ID NO:95 and, preferably, a leucine-rich-repeat domain is located at about amino acid positions 26 to 57, 58 to 81, 82 to 105, 106 to 130, 131 to 154, 155 to 178, 179 to 202, 203 to 226, 227 to 250, and/or 260 to 310 of mouse TANGO 393 (SEQ ID NO:95), SEQ ID NO:99, 100, 101, 102, 103, 104, 105, 106, 107 and 108, respectively.

In another embodiment, a TANGO 393 family member includes one or more leucine-rich-repeat domain consensus sequences having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 26 to 57, 58 to 81, 82 to 105, 106 to 130, 131 to 154, 155 to 178, 179 to 202, 203 to 226, 227 to 250, and/or 260 to 310 of human TANGO 393 of SEQ ID NO:76 (SEQ ID NO:79, 80, 81, 83, 83, 84, 85, 86, 87 and 88, respectively). In another embodiment, a TANGO 393 family member includes one or more leucine-rich-repeat domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acid positions 26 to 57, 58 to 81, 82 to 105, 106 to 130, 131 to 154, 155 to 178, 179 to 202, 203 to 226, 227 to 250, and/or 260 to 310 of mouse TANGO 393 (SEQ ID NO:95), SEQ ID NO:99, 100, 101, 102, 103, 104, 105, 106, 107 and 108, respectively.

In another embodiment, a TANGO 393 family member includes one or more leucine-rich-repeat domain consensus sequences having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 26 to 57, 58 to 81, 82 to 105, 106 to 130, 131 to 154, 155 to 178, 179 to 202, 203 to 226, 227 to 250, and/or 260 to 310 of human TANGO 393 of SEQ ID NO:76 (SEQ ID NO:79, 80, 81, 83, 83, 84, 85, 86, 87 and 88, respectively), and has at least one TANGO 393 biological activity as described herein. In yet another embodiment, a TANGO 393 family member includes one or more leucine-rich-repeat domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acid positions 26 to 57, 58 to 81, 82 to 105, 106 to 130, 131 to 154, 155 to 178, 179 to 202, 203 to 226, 227 to 250, and/or 260 to 310 of mouse TANGO 393 (SEQ ID NO:95), SEQ ID NO:99, 100, 101, 102, 103, 104, 105, 106, 107 and 108, respectively, and has at least one TANGO 393 biological activity as described herein.

A TANGO 402 family member can include of one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain. In one embodiment, a TANGO 402 protein contains an extracellular domain at amino acids 1 to about 133 of SEQ ID NO:112 or a mature extracellular domain at about amino acid residues 51 to 133 of SEQ ID NO:112 (SEQ ID NO:115), a transmembrane domain at about amino acid residues 134 to 151 of SEQ ID NO:112 (SEQ ID NO:116),

and a cytoplasmic domain at about amino acid residues 152 to 207 of SEQ ID NO:112 (SEQ ID NO:117).

A TANGO 402 family member can include a signal sequence. In certain embodiments, a TANGO 402 family member has the amino acid sequence of SEQ ID NO:112, and the signal sequence is located at amino acids 1 to 48, 1 to 49, 1 to 50, 1 to 51 or 1 to 52. In such embodiments of the invention, the extracellular domain and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 48 results in an extracellular domain consisting of amino acids 49 to 133 of SEQ ID NO:112 and the mature TANGO 402 protein corresponding to amino 49 to 207 of SEQ ID NO:112.

A TANGO 402 family member can include a C-type lectin domain or a C-type lectin-like domain.

A C-type lectin domain typically has the following consensus sequence: C-[LIVMFATG]-x(5,12)-[WL]-x-[DNSR]-x(2)-C-x(5,6)-[FYWLIVSTA]-[LIVSTA]-C, wherein C is a cysteine residue, [LIVMFATG] is a leucine, isoleucine, methionine, phenylalanine, alanine, threonine or glycine residue, x is any amino acid and the number in parentheses indicates the number of amino acids, [WL] is either a tryptophan or leucine residue, [DNSR] is a aspartic acid, asparagine, serine or arginine residue, [FYWLIVSTA] is a phenylalanine, tyrosine, tryptophan, leucine, isoleucine, valine, serine, threonine or alanine residue, and [LIVSTA] is a leucine, isoleucine, valine, serine, threonine or alanine residue. C-type lectin domains contain four cysteines, which are involved in two disulfide bonds, and are about 110 to 130 amino acid residues. C-type lectin domains typically function as calcium-dependent carbohydrate-recognition domains and have been found in various proteins including, but not limited to, asialoglycoprotein receptors (ASGPR), pulmonary surfactant-associated protein A (SP-A), mannan-binding proteins, L-selectin, neurocan, and tetranectin. These proteins have various functions including, for example, cell adhesion (*i.e.*, L-selectin). ASGPR mediates the endocytosis of plasma glycoprotein to which the terminal salic acid-residue in their carbohydrated moieties has been removed. SP-A binds to surfactant phospholipids and contributes to lower the surface tension at the air-liquid interface in the alveoli of the lung.

A C-type lectin-like domain as described herein has the following consensus sequence: C-[LIVMFATG]-x-(5,12)-[DNSR]-x(2)-C-x(5,6)-[LIVSTA]-C, wherein C is a cysteine residue, [LIVMFATG] is a leucine, isoleucine, methionine, phenylalanine, alanine, threonine or glycine residue, "x" is any amino acid and the number in parentheses indicates the number of amino acids [DNSR] is an aspartic acid, asparagine, serine or arginine residue, and [LIVSTA] is a leucine, isoleucine, valine, serine, threonine or

alanine residue. In one embodiment, a TANGO 402 family member has the amino acid sequence of SEQ ID NO:112 and, preferably, a C-type lectin-like domain is located at about amino acid positions 104 to 193 of SEQ ID NO:112 (SEQ ID NO:118), wherein the consensus sequence is at about amino acid positions 172 to 193 of SEQ ID NO:112 (SEQ ID NO:119).

In another embodiment, a TANGO 402 family member includes one or more C-type lectin-like domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 104 to 193 of SEQ ID NO:112 (SEQ ID NO:118).

In another embodiment, a TANGO 402 family member includes one or more C-type lectin-like domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 104 to 193 of SEQ ID NO:112 (SEQ ID NO:118), and has at least one TANGO 402 biological activity as described herein.

In another embodiment, a TANGO 402 family member includes one or more C-type lectin-like domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 104 to 193 of SEQ ID NO:112 (SEQ ID NO:118) and includes a cysteine residue N-terminal to the consensus sequence. In yet another embodiment, a TANGO 402 family member includes one or more C-type lectin-like domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 104 to 193 of SEQ ID NO:112 (SEQ ID NO:118), includes a cysteine residue N-terminal to the consensus sequence, and has at least one TANGO 402 biological activity as described herein.

In another embodiment, the C-type lectin-like domain of TANGO 402 is a C-type lectin domain, which has the following consensus sequence: C-[LIVMFATG]-x(5,12)-[WL]-x-[DNSR]-x(2)-C-x(5,6)-[FYWLIVSTA]-[LIVSTA]-C, wherein C is a cysteine residue, [LIVMFATG] is a leucine, isoleucine, methionine, phenylalanine, alanine, threonine or glycine residue, x is any amino acid, [WL] is either a tryptophan or leucine residue, [DNSR] is a aspartic acid, asparagine, serine or arginine residue, [FYWLIVSTA] is a phenylalanine, tyrosine, tryptophan, leucine, isoleucine, valine, serine, threonine or alanine residue, and [LIVSTA] is a leucine, isoleucine, valine, serine, threonine or alanine

residue. In this embodiment, a TANGO 402 family member includes one or more C-type lectin-like domains having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 104 to 193 of SEQ ID NO:112 (SEQ ID NO:118).

A MANGO 346 family member can include a signal sequence. In certain embodiments, a MANGO 346 family member has the amino acid sequence of SEQ ID NO:125, and the signal sequence is located at amino acids 1 to 17, 1 to 18, 1 to 19, 1 to 20 or 1 to 21. In such embodiments of the invention, the extracellular domain and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 19 results in the mature MANGO 346 protein corresponding to amino 20 to 60 (SEQ ID NO:127).

A MANGO 349 family member can include a signal sequence. In certain embodiments, a MANGO 349 family member has the amino acid sequence of SEQ ID NO:130, and the signal sequence is located at amino acids 1 to 24, 1 to 25, 1 to 26, 1 to 27 or 1 to 28. In such embodiments of the invention, the extracellular domain and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 26 results in the mature MANGO 349 protein corresponding to amino 27 to 167 of SEQ ID NO:130 (SEQ ID NO:132).

Various features of TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 are summarized below.

<u>Human TANGO 339</u>

A cDNA encoding human TANGO 339 was identified by analyzing the sequences of clones present in a human fetal library for sequences that encode wholly secreted or transmembrane proteins. This analysis led to the identification of a clone, jthga100g01, encoding full-length human TANGO 339. The human TANGO 339 cDNA of this clone is 2715 nucleotides long (Figure 1; SEQ ID NO:1). The open reading frame of this cDNA, nucleotides 210 to 1019 of SEQ ID NO:1 (SEQ ID NO:2), encodes a 270 amino acid transmembrane protein (Figure 1; SEQ ID NO:3).

Figure 2 depicts a hydropathy plot of human TANGO 339. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) and N-glycosylation site are indicated by short vertical lines just below the

hydropathy trace. The dashed vertical line separates the signal sequence on the left from the mature protein on the right.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, Protein Engineering 10:1-6) predicted that human TANGO 339 includes a 42 amino acid signal peptide (amino acid 1 to amino acid 42 of SEQ ID NO:3; SEQ ID NO:5) preceding the mature human TANGO 339 protein (corresponding to amino acid 43 to amino acid 270 of SEQ ID NO:3; SEQ ID NO:4). In instances wherein the signal peptide is cleaved, the molecular weight of human TANGO 339 protein without post-translational modifications is 30.7 kDa prior to the cleavage of the signal peptide, and 25.6 kDa after cleavage of the signal peptide. The presence of a methionine residue at positions 56, 67 and 72 of SEQ ID NO:3 indicates that there can be alternative forms of human TANGO 339 of 215 amino acids of SEQ ID NO:3, 204 amino acids of SEQ ID NO:3, and 199 amino acids of SEQ ID NO:3, respectively.

Human TANGO 339 protein is a transmembrane protein that contains extracellular domains at amino acid residues 43 to 61 and 116 to 232 of SEQ ID NO:3 (SEQ ID NO:20 and SEQ ID NO:21, respectively), transmembrane domains at amino acid residues 62 to 84, 93 to 115, and 233 to 254 of SEQ ID NO:3 (SEQ ID NO:15, SEQ ID NO:16 and SEQ ID NO:17, respectively), and cytoplasmic domains at amino acid residues 85 to 92 and 255 to 270 of SEQ ID NO:3 (SEQ ID NO:22 and SEQ ID NO:23, respectively).

In instances wherein the signal peptide is not cleaved, human TANGO 339 has extracellular domains at amino acid residues 1 to 16, 85 to 92, and 255 to 270 of SEQ ID NO:3 (SEQ ID NO:11, SEQ ID NO:12, and SEQ ID NO:13, respectively), transmembrane domains at amino acid residues 17 to 41, 62 to 84, 93 to 115, and 233 to 254 of SEQ ID NO:3 (SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, and SEQ ID NO:17, respectively), and cytoplasmic domains of amino acid residues 42 to 61 and 116 to 232 of SEQ ID NO:3 (SEQ ID NO:18 and SEQ ID NO:19, respectively).

Alternatively, in another embodiment, a human TANGO 339 protein contains cytoplasmic domains at amino acid residues 43 to 61 and 116 to 232 of SEQ ID NO:3, transmembrane domains at amino acid residues 62 to 84, 93 to 115, and 233 to 254 of SEQ ID NO:3 (SEQ ID NO:15, SEQ ID NO:16 and SEQ ID NO:17, respectively), and extracellular domains at amino acid residues 85 to 92 and 255 to 270 of SEQ ID NO:3.

In one embodiment of a nucleotide sequence of human TANGO 339, the nucleotide at position 29 is adenine (A)(SEQ ID NO:2). In this embodiment, the amino acid at position 10 is lysine (K)(SEQ ID NO:3). In an alternative embodiment, a species variant of human TANGO 339 has a nucleotide at position 29 which is guanine (G)(SEQ

ID NO:136). In this embodiment, the amino acid at position 10 is arginine (R)(SEQ ID NO:137), *i.e.*, a conservative substitution.

In another embodiment of a nucleotide sequence of human TANGO 339, the nucleotide at position 59 is thymine (T)(SEQ ID NO:2). In this embodiment, the amino acid at position 20 is phenylalanine (F)(SEQ ID NO:3). In an alternative embodiment, a species variant of human TANGO 339 has a nucleotide at position 59 which is adenine (A)(SEQ ID NO:138). In this embodiment, the amino acid at position 20 is tyrosine (Y)(SEQ ID NO:139), *i.e.*, a conservative substitution.

In another embodiment of a nucleotide sequence of human TANGO 339, the nucleotide at position 119 is cytosine (C)(SEQ ID NO:2). In this embodiment, the amino acid at position 40 is alanine (A)(SEQ ID NO:3). In an alternative embodiment, a species variant of human TANGO 339 has a nucleotide at position 119 which is thymine (T)(SEQ ID NO:140). In this embodiment, the amino acid at position 40 is valine (V)(SEQ ID NO:141), *i.e.*, a conservative substitution.

In another embodiment of a nucleotide sequence of human TANGO 339, the nucleotide at position 180 is cytosine (C)(SEQ ID NO:2). In this embodiment, the amino acid at position 60 is aspartate (D)(SEQ ID NO:3). In an alternative embodiment, a species variant of human TANGO 339 has a nucleotide at position 180 which is guanine (G)(SEQ ID NO:142). In this embodiment, the amino acid at position 60 is glutamate (E)(SEQ ID NO:143), *i.e.*, a conservative substitution.

Human TANGO 339 includes a transmembrane 4-like domain (at amino acids 68 to 260 of SEQ ID NO:3; SEQ ID NO:6) and a peripherin /rom-1-like domain (at amino acids 18 to 270 of SEQ ID NO:3; SEQ ID NO:8).

Human TANGO 339 has an N-glycosylation site with the sequence NCSG (at amino acid residues 169 to 172 of SEQ ID NO:3). Two protein kinase C phosphorylation sites are present in human TANGO 339. The first has the sequence SEK (at amino acid residues 42 to 44 of SEQ ID NO:3) and the second has the sequence SYR (at amino acid residues 133 to 135 of SEQ ID NO:3). Human TANGO 339 has three casein kinase II phosphorylation sites. The first has the sequence SYRD (at amino acid residues 133 to 136 of SEQ ID NO:3), the second has the sequence SKWD (at amino acid residues 210 to 213 of SEQ ID NO:3), and the third has the sequence SDIE (at amino acid residues 259 to 262 of SEQ ID NO:3). Six N-myristylation sites are present in human TANGO 339. The first has the sequence GCVGAL (at amino acid residues 79 to 84 of SEQ ID NO:3), the second has the sequence GASYSR (at amino acid residues 172 to 177 of SEQ ID NO:3), the third has the sequence GVPFSC (at amino acid residues 181 to 186 of SEQ ID NO:3), the fourth has the sequence GCIQAL (at amino acid residues 220 to 225 of SEQ ID

NO:3), the fifth has the sequence GVFIAI (at amino acid residues 238 to 243 of SEQ ID NO:3), and the sixth has the sequence GIFLAR (at amino acid residues 250 to 255 of SEQ ID NO:3). Human TANGO 339 has a prokaryotic membrane lipoprotein lipid attachment site with the sequence VVMFTLGFAGC (at amino acid residues 70 to 80 of SEQ ID NO:3; SEQ ID NO:10).

The human TANGO 339 gene maps to human chromosome 10 between markers D10S201 and D10S551. As retinal G protein coupled receptor and pulmonary-associated protein A1 map to this region of chromosome 10, TANGO 339 nucleic acids, proteins and modulators thereof can be used to diagnose disorders associated with G protein coupled receptors and/or modulate G protein coupled receptor-related processes, *e.g.*, retinal processes and/or pulmonary-related processes.

Figure 3 shows an alignment of the human TANGO 339 amino acid sequence (SEQ ID NO:3) with the human CD9 antigen amino acid sequence (SEQ ID NO:24; Accession Number NM_001769). The alignment shows that there is a 24.1% overall amino acid sequence identity between human TANGO 339 and human CD9 antigen. The CD9 antigen is a widely expressed cell surface glycoprotein that has been shown to be involved in such processes as cell activation, proliferation, and adhesion. For example, CD9 antigen expression on platelets mediates platelet activation and aggregation. CD9 antigen has also been shown to be expressed by neural cells and can play a role in intercellular signaling in the nervous system, in particular, controlling cellular attraction or repulsion in guiding neural growth to target points. Further, the CD9 antigen has been shown to associate with beta 1 integrins and other transmembrane 4 superfamily members, including CD81 and CD82. As such TANGO 339 proteins, nucleic acids and modulators thereof could be useful in modulating cellular interaction such as between immune cells, and also can be involved in modulating intercellular signaling, such as neural cell intercellular signaling.

Figure 4 shows an alignment of the nucleotide sequence of human CD9 antigen coding region (SEQ ID NO:25; Accession Number NM_001769) and the nucleotide sequence of human TANGO 339 coding region (SEQ ID NO:2). The alignment shows a 45.9 % overall sequence identity between the two nucleotide sequences. The full-length human CD9 antigen nucleic acid sequence (SEQ ID NO:26; Accession Number NP_001760) and human TANGO 339 cDNA (SEQ ID NO:1) have an overall sequence identity of 30.3%.

Clone EpT339, which encodes human TANGO 339, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-292. This deposit will be

maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Uses of TANGO 339 Nucleic acids, Polypeptides, and Modulators Thereof

As TANGO 339 was originally found in a human fetal library, TANGO 339 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, development, differentiation, and/or function of cells, tissues and/or organs, *e.g.*, the proliferation of tissues and cells of lymphoid origin and neural origin. TANGO 339 nucleic acids, proteins and modulators thereof can be used to treat immune related disorders, *e.g.*, immunodeficiency disorders (*e.g.*, HIV), viral disorders, cancers, autoimmune disorders, (*e.g.*, arthritis and graft rejection) and inflammatory disorders (*e.g.*, bacterial or viral infection, psoriasis, septicemia, arthritis, allergic reactions). TANGO 339 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the development of cells, tissues and/or organs in the embryo and/or fetus.

In light of the fact that TANGO 339 has characteristics of transmembrane 4 proteins, TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate (*e.g.*, stabilize, promote, inhibit or disrupt) cellular activation, cellular proliferation, motility, and differentiation. For example, such TANGO 339 compositions and modulators thereof can be used to modulate binding to extracellular matrix (ECM)-associated factors such as integrins and can function to modulate ligand binding to cell surface receptors.

In further light of the fact that TANGO 339 has characteristics of transmembrane 4 proteins, TANGO 339 nucleic acids, proteins and modulators thereof can be used to modulate disorders associated with aberrant signal transduction in response to ECM-associated proteins and cell surface receptors such as other transmembrane 4 proteins. TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate the development and progression of proliferative disorders, *e.g.*, neoplasms or tumors (such as carcinomas, sarcomas, adenomas or myeloid lymphomas) associated with cancer, (*e.g.*, fibrosarcoma, myxosarcoma, liposarcoma, chondrosarcoma, osteogenic sarcoma, chordoma, angiosarcoma, endotheliosarcoma, lymphangiosarcoma, lymphangioendotheliosarcoma, synovioma, mesothelioma, Ewing's tumor, leimyosarcoma, rhabdotheliosarcoma, colon sarcoma, pancreatic cancer, breast cancer, ovarian cancer, prostate cancer, squamous cell carcinoma, basal cell carcinoma,

adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinomas, cystadenocarcinoma, medullary carcinoma, bronchogenic carcinoma, renal cell carcinoma, hematoma, bile duct carcinoma, melanoma, choriocarcinoma, semicoma, embryonal carcinoma, Wilms' tumor, cervical cancer, testicular tumor, lung carcinoma, small cell carcinoma, bladder carcinoma, epithelial carcinoma, glioma, astrocytoma, medulloblastoma, craniopharyngioma, ependynoma, pinealoma, hemangioblastoma, retinoblastoma; leukemias, *e.g.* acute lymphocytic leukemia and acute myelocytic leukemia (myelolastic, promyelocytic, myelomonocytic, monocytic and erythroleukemia); chronic leukemia (chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia); and polycythemia vera, lymphoma (Hodgkin's disease and non-Hodgkin's diseases), multiple myeloma and Waldenström's macroglobulinemia.

TANGO 339 proteins exhibit similarity to human CD9 antigen, a member of the transmembrane 4 superfamily. In light of this, TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate platelet activation and aggregation. For example, antagonists to TANGO 339 action, such as peptides, antibodies or small molecules that decrease or block TANGO 339 binding to extracellular matrix components (*e.g.*, integrins) or that prevent TANGO 339 signaling, can be used as platelet activation and aggregation blockers. In another example, agonists that mimic TANGO 339 activity, such as peptides, antibodies or small molecules, can be used to induce platelet activation and aggregation. Antibodies may activate or inhibit the cell adhesion, proliferation and activation, and may help in treating inflammation, cancer, cardiovascular disease or stroke by affecting these cellular processes. TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate platelet-related processes and disorders, *e.g.*, Glanzmann's thromboasthemia, which is a bleeding disorder characterized by failure of platelet aggregation in response to cell stimuli.

In further light of the fact that TANGO 339 proteins exhibit similarity to human CD9 antigen, TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate intercellular signaling in the nervous system. The CD9 antigen, which is expressed at the surface of central nervous system (CNS) mature myelin, may modulate intercellular signal transduction and enhance myelin membrane adhesion to extracellular matrices at very late stages of development, thereby playing a role in the maintenance of the entire myelin sheath.

In light, in part, of the fact that TANGO 339 proteins contain peripherin/rom-1-like domains, TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate the development and function of the eye, such as retinal development and

function, (*e.g.*, photoreceptor disk morphogenesis). TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to treat eye diseases and/or disorders, *e.g.*, autosomal dominant retinitis pigmentosa, autosomal dominant punctata albescens, butterfly-shaped pigment dystrophy, cataracts, macular degeneration, myopia, stigmatism and retinoblastoma.

As TANGO 339 maps to a region of chromosome 10 which encodes polypeptides expressed in the lung, TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate the development, differentiation and activity of pulmonary structures, *e.g.*, lung. TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate or treat pulmonary disorders, such as atelectasis, pulmonary congestion or edema, cystic fibrosis, chronic obstructive airway disease (*e.g.*, emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (*e.g.*, sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), lung cancer or tumors (*e.g.*, bronchogenic carcinoma, bronchiolovlveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors).

As TANGO 339 nucleic acids exhibit homology to a human brain EST (Accession Number Q59384, disclosed in Patent No. WP 93/16178), TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to modulate processes involved in the development, differentiation and activity of the brain, including, but not limited to development, differentiation and activation of neuronal cells and glial cells (*e.g.*, oligodendrocytes astrocytes), and amelioration of one or more symptoms associated with abnormal function of such cell types. TANGO 339 nucleic acids, proteins and modulators thereof can be utilized to treat neural diseases and/or disorders, *e.g.* epilepsy, spinal cord injuries, infarction, infection, malignancy, paraneoplastic syndromes, neuropsychiatric disorders (*e.g.*, schizophrenia, depression, anxiety disorders, and anorexia nervosa), and neurodegenerative diseases including, but not limited to, Alzheimer's disease, Parkinson's disease, Huntington's Chorea, amyotrophic lateral sclerosis and progressive supra-nuclear palsy.

TANGO 339 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the brain) and/or cells (*e.g.*, neurons) in which TANGO 339 is expressed. TANGO 339 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

Human TANGO 353

A cDNA encoding human TANGO 353 was identified by analyzing the sequences of clones present in a mixed lymphocyte reaction library for sequences that encode a wholly secreted or transmembrane protein. This analysis led to the identification of a clone, jthLa031g12 encoding full-length human TANGO 353. The human TANGO 353 cDNA of this clone is 1239 nucleotides long (Figure 5; SEQ ID NO:27). The open reading frame of this cDNA, nucleotides 76 to 765 of SEQ ID NO:27 (SEQ ID NO:28), encodes a 230 amino acid transmembrane protein (Figure 5; SEQ ID NO:29).

Figure 6 depicts a hydropathy plot of human TANGO 353. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) and N-glycosylation sites (NGly) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence on the left from the mature protein on the right.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, Protein Engineering 10:1-6) predicted that human TANGO 353 includes a 14 amino acid signal peptide (amino acid 1 to amino acid 14 of SEQ ID NO:29; SEQ ID NO:31) preceding the mature human TANGO 353 protein (corresponding to amino acid 15 to amino acid 230 of SEQ ID NO:29; SEQ ID NO:30). The molecular weight of human TANGO 353 protein without post-translational modifications is 24.8 kDa prior to the cleavage of the signal peptide and 23.3 kDa after cleavage of the signal peptide. The presence of a methionine residue at positions 39, 170 and 184 of SEQ ID NO:29 indicates that there can be alternative forms of human TANGO 353 of 192 amino acids of SEQ ID NO:29, 61 amino acids of SEQ ID NO:29, and 47 amino acids of SEQ ID NO:29, respectively.

Human TANGO 353 is a transmembrane protein which can include one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain. The human TANGO 353 protein contains an extracellular domain at amino acid residues 15 to 116 of SEQ ID NO:29 (SEQ ID NO:32), a transmembrane domain at amino acid residues 117 to 141 of SEQ ID NO:29 (SEQ ID NO:33), and a cytoplasmic domain at amino acid residues 142 to 230 of SEQ ID NO:29 (SEQ ID NO:34).

Alternatively, in another embodiment, a human TANGO 353 protein contains a cytoplasmic domain at amino acid residues 15 to 116 of SEQ ID NO:29, a transmembrane domain at amino acid residues 117 to 141 of SEQ ID NO:29 (SEQ ID NO:33), and an extracellular domain at amino acid residues 142 to 230 of SEQ ID NO:29.

In one embodiment of a nucleotide sequence of human TANGO 353, the nucleotide at position 68 is thymine (T)(SEQ ID NO:28). In this embodiment, the amino acid at position 23 is valine (V)(SEQ ID NO:29). In an alternative embodiment, a species variant of human TANGO 353 has a nucleotide at position 68 which is cytosine (C)(SEQ ID NO:144). In this embodiment, the amino acid at position 23 is alanine (A)(SEQ ID NO:145), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 353, the nucleotide at position 77 is adenine (A)(SEQ ID NO:28). In this embodiment, the amino acid at position 26 is tyrosine (Y)(SEQ ID NO:29). In an alternative embodiment, a species variant of human TANGO 353 has a nucleotide at position 77 which is thymine (T)(SEQ ID NO:146). In this embodiment, the amino acid at position 26 is phenylalanine (F)(SEQ ID NO:147), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 353,.the nucleotide at position 203 is guanine (G)(SEQ ID NO:28). In this embodiment, the amino acid at position 68 is arginine (R)(SEQ ID NO:29). In an alternative embodiment, a species variant of human TANGO 353 has a nucleotide at position 203 which is adenine (A)(SEQ ID NO:148). In this embodiment, the amino acid at position 68 is histidine (H)(SEQ ID NO:149), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 353, the nucleotide at position 309 is cytosine (C)(SEQ ID NO:28). In this embodiment, the amino acid at position 103 is aspartate (D)(SEQ ID NO:29). In an alternative embodiment, a species variant of human TANGO 353 has a nucleotide at position 309 which is guanine (G)(SEQ ID NO:150). In this embodiment, the amino acid at position 103 is glutamate (E)(SEQ ID NO:151), *i.e.*, a conservative substitution.

Four N-glycosylation sites are present in human TANGO 353. The first has the sequence NFTL (at amino acid residues 48 to 51 of SEQ ID NO:29), the second has the sequence NLSG (at amino acid residues 73 to 76 of SEQ ID NO:29), the third has the sequence NQSQ (at amino acid residues 97 to 100 of SEQ ID NO:29), and the fourth has the sequence NVSF (at amino acid residues 109 to 112 of SEQ ID NO:29). Human TANGO 353 has one cAMP- and cGMP-dependent protein kinase phosphorylation site with the sequence KRPT (at amino acid residues 209 to 212 of SEQ ID NO:29). Five protein kinase C phosphorylation sites are present in human TANGO 353. The first has the sequence SIR (at amino acid residues 19 to 21 of SEQ ID NO:29), the second has the sequence SSK (at amino acid residues 78 to 80 of SEQ ID NO:29), the third has the sequence SAK (at amino acids 180 to 182 of SEQ ID NO:29), the fourth has the sequence TRK (at amino acid residues 207 to 209 of SEQ ID NO:29), and the fifth has the sequence

TFR (at amino acid residues 225 to 227 of SEQ ID NO:29). Human TANGO 353 has four casein kinase II phosphorylation sites. The first has the sequence SSQE (at amino acid residues 28 to 31 of SEQ ID NO:29), the second has the sequence TMPE (at amino acid residues 183 to 186 of SEQ ID NO:29), the third has the sequence TLDD (at amino acid residues 191 to 194 of SEQ ID NO:29), and the fourth has the sequence SSPE (at amino acid residues 216 to 219 of SEQ ID NO:29). Human TANGO 353 has two N-myristylation sites. The first has the sequence GNFPGA (at amino acid residues 42 to 47 of SEQ ID NO:29) and the second has the sequence GVTFNL (at amino acid residues 69 to 74 of SEQ ID NO:29).

Clone EpT353, which encodes human TANGO 353, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-292. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Uses of TANGO 353 Nucleic acids, Polypeptides, and Modulators Thereof

As TANGO 353 was originally found in a mixed lymphocyte library, TANGO 353 nucleic acids, proteins, and modulators thereof can be utilized to diagnose disorders and/or modulate processes involved in lymphocyte development, differentiation and activity, including, but not limited to development, differentiation and activation of T cells, including T helper, T cytotoxic and non-specific T killer cell types and subtypes, and B cells, immune functions associated with such cells, and amelioration of one or more symptoms associated with abnormal function of such cell types. Such disorders can include, but are not limited to, autoimmune disorders (*e.g.*, autoimmune thyroiditis, Type I diabetes mellitus, insulin-resistant diabetes, autoimmune anemia, multiple sclerosis, rheumatoid arthritis, lupus or sclerodoma, allergy, including allergic rhinitis and food allergies, asthma, psoriasis, graft rejection, transplantation rejection, graft versus host disease, pathogenic susceptibilities), inflammatory disorders (*e.g.*, bacterial or viral infections, wound healing and inflammatory bowel disease and arthritis), apoptotic disorders, and cytotoxic disorders, septic shock, cachexia, and proliferative disorders (*e.g.*, B cell cancers stimulated by TNF).

Other TANGO 353 associated disorders can include TNF related disorders (*e.g.*, acute myocarditis, myocardial infarction, congestive heart failure, T cell disorders (*e.g.*, dermatitis, fibrosis)), immunological differentiative and apoptotic disorders (*e.g.*, hyper-

proliferative syndromes such as systemic lupus erythematosus (lupus)), and disorders related to angiogenesis (*e.g.*, tumor formation and/or metastasis, cancer). Modulators of TANGO 353 expression and/or activity can be used to treat such disorders.

As TANGO 353 is expressed in mixed lymphocyte cultures, and hence likely expressed in bone marrow, TANGO 353 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders associated with cells in the bone marrow and/or modulate the proliferation, differentiation, and/or function of cells that appear in the bone marrow, *e.g.*, stem cells (*e.g.*, hematopoietic stem cells), and blood cells, *e.g.*, erythrocytes, platelets, and leukocytes. Thus TANGO 353 nucleic acids, proteins, and modulators thereof can be used to treat bone marrow, blood, and hematopoietic associated diseases and disorders, *e.g.*, acute myeloid leukemia, hemophilia, leukemia, anemia (*e.g.*, sickle cell anemia), and thalassemia.

As TANGO 353 is a transmembrane protein, TANGO 353 nucleic acids, proteins and modulators thereof can be utilized to modulate intercellular signaling cascades, or alternatively.

TANGO 353 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, spleen) and/or cells (*e.g.*, lymphocytes) in which TANGO 353 is expressed. TANGO 353 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

<u>Human TANGO 358</u>

A cDNA encoding human TANGO 358 was identified by analyzing the sequences of clones present in a fetal thymus library for sequences that encode a wholly secreted or transmembrane protein. This analysis led to the identification of a clone, jthTb128c07 encoding full-length human TANGO 358. The human TANGO 358 cDNA of this clone is 1608 nucleotides long (Figure 7; SEQ ID NO:36). The open reading frame of this cDNA, nucleotides 184 to 429 of SEQ ID NO:36 (SEQ ID NO:37), encodes an 82 amino acid transmembrane protein (Figure 7; SEQ ID NO:38).

Figure 8 depicts a hydropathy plot of human TANGO 358. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence on the left from the mature protein on the right.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, Protein Engineering <u>10</u>:1-6) predicted that human TANGO 358 includes a 42 amino acid signal peptide (amino acid 1 to amino acid 42 of SEQ ID NO:36; SEQ ID NO:40) preceding the

mature human TANGO 358 protein (corresponding to amino acid 43 to amino acid 82 of SEQ ID NO:36; SEQ ID NO:39). The molecular weight of human TANGO 358 protein without post-translational modifications is 9.5 kDa prior to the cleavage of the signal peptide and 4.5 kDa after cleavage of the signal peptide. The presence of a methionine residue at positions 17, 20 and 63 of SEQ ID NO:38 indicates that there can be alternative forms of human TANGO 358 of 66 amino acids of SEQ ID NO:38, 63 amino acids of SEQ ID NO:38, and 20 amino acids of SEQ ID NO:38, respectively.

Human TANGO 358 is a transmembrane protein which can include one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain. The human TANGO 358 protein contains an extracellular domain at amino acid residues 43 to 49 of SEQ ID NO:38 (SEQ ID NO:41), a transmembrane domain at amino acid residues 50 to 66 of SEQ ID NO:38 (SEQ ID NO:42), and a cytoplasmic domain at amino acid residues 67 to 82 of SEQ ID NO:38 (SEQ ID NO:43).

Alternatively, in another embodiment, a human TANGO 358 protein contains a cytoplasmic domain at amino acid residues 43 to 49 of SEQ ID NO:38, a transmembrane domain at amino acid residues 50 to 66 of SEQ ID NO:38 (SEQ ID NO:42), and an extracellular domain at amino acid residues 67 to 82 of SEQ ID NO:38. Further, human TANGO 358 has a protein kinase C phosphorylation site with the sequence SIK (at amino acid residues 45 to 47 of SEQ ID NO:38).

In one embodiment of a nucleotide sequence of human TANGO 358, the nucleotide at position 20 is adenine (A)(SEQ ID NO:37). In this embodiment, the amino acid at position 7 is histidine (H)(SEQ ID NO:38). In an alternative embodiment, a species variant of human TANGO 358 has a nucleotide at position 20 which is guanine (G)(SEQ ID NO:152). In this embodiment, the amino acid at position 7 is arginine (R)(SEQ ID NO:153), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 358, the nucleotide at position 35 is thymine (T)(SEQ ID NO:37). In this embodiment, the amino acid at position 12 is valine (V)(SEQ ID NO:38). In an alternative embodiment, a species variant of human TANGO 358 has a nucleotide at position 35 which is cytosine (C)(SEQ ID NO:154). In this embodiment, the amino acid at position 12 is alanine (A)(SEQ ID NO:155), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 358, the nucleotide at position 85 is thymine (T)(SEQ ID NO:37). In this embodiment, the amino acid at position 29 is serine (S)(SEQ ID NO:38). In an alternative embodiment, a species variant of human TANGO 358 has a nucleotide at position 85 which is adenine (A)(SEQ

ID NO:156). In this embodiment, the amino acid at position 29 is threonine (T)(SEQ ID NO:157), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 358, the nucleotide at position 91 is cytosine (C)(SEQ ID NO:37). In this embodiment, the amino acid at position 31 is glutamine (Q)(SEQ ID NO:38). In an alternative embodiment, a species variant of human TANGO 358 has a nucleotide at position 91 which is guanine (G)(SEQ ID NO:158). In this embodiment, the amino acid at position 31 is glutamate (E)(SEQ ID NO:159), *i.e.*, a conservative substitution.

Clone EpT358, which encodes human TANGO 358, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999, and assigned Accession Number PTA-292. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Uses of TANGO 358 Nucleic acids, Polypeptides, and Modulators Thereof

As TANGO 358 was originally found in a fetal thymus library, TANGO 358 nucleic acids, proteins, and modulators thereof can be used to diagnose thymus associated disorders. TANGO 358 nucleic acids, proteins, and modulators thereof can also be used modulate the proliferation, development, differentiation, maturation and/or function of thymocytes, *e.g.*, modulate development and maturation of T-lymphocytes. TANGO 358 nucleic acids, proteins and modulators thereof can be utilized to modulate immune-related processes such as the ability to modulate host immune response by, *e.g.*, modulating the formation of and/or binding to immune complexes, and modulating the positive and negative selection of thymocytes. Such TANGO 358 compositions and modulators thereof can be utilized, *e.g.*, to ameliorate incidence of any symptoms associated with disorders that involve such immune-related processes, including, but not limited to infection and autoimmune disorders (*e.g.*, insulin-dependent mellitus, multiple sclerosis, systemic lupus, erythematosus, sjogren's syndrome, autoimmune thyroiditis, idiotpathic Addison's disease, vitiligo, Grave's disease, idiopathic thrombocytopenia purpura, rheumatoid arthritis, and scleroderma). TANGO 358 nucleic acids, proteins and modulators thereof can also be utilized to treat viral infections, inflammatory immune disorders and immune-related cancers including but not limited to, leukemia (*e.g.*, acute leukemia, chronic leukemia, Hodgkin's disease non-Hodgkin's lymphoma ,and multiple myeloma).

Disorders associated with TANGO 358 activity, including those which TANGO 358 proteins, nucleic acids and modulators thereof may be an antagonist can be used to treat include immune disorders, *e.g.*, autoimmune disorders (*e.g.*, arthritis, graft rejection (*e.g.*, allograft rejection), T cell disorders (*e.g.*, AIDS)) and inflammatory disorders (*e.g.*, bacterial infection, psoriasis, septicemia, cerebral malaria, inflammatory bowel disease, arthritis (*e.g.*, rheumatoid arthritis, osteoarthritis), and allergic inflammatory disorders (*e.g.*, asthma, psoriasis)). Disorders associated with modulated TANGO 358 activity can also include apoptotic disorders (*e.g.*, rheumatoid arthritis, systemic lupus erythematosus, insulin-dependent diabetes mellitus), cytotoxic disorders, septic shock, cachexia, and proliferative disorders (*e.g.*, B cell cancers stimulated by TNF).

In light of the fact that TANGO 358 was isolated from a thymus library, TANGO 358 proteins, nucleic acids and modulators thereof can be used to treat disorders that include TNF-related disorders (*e.g.*, acute myocarditis, myocardial infarction, congestive heart failure, T cell disorders (*e.g.*, dermatitis, fibrosis)), differentiative and apoptotic disorders, and disorders related to angiogenesis (*e.g.*, tumor formation and/or metastasis, cancer). Modulators of TANGO 358 expression and/or activity can be used to treat such disorders.

As TANGO 358 is a transmembrane protein, TANGO 358 nucleic acids, proteins and modulators thereof can be utilized to diagnose disorders and/or modulate intercellular signaling pathways, for example by disrupting ligand-receptor interactions or cellular interactions with the extra-cellular matrix.

TANGO 358 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the thymus) and/or cells (*e.g.*, T-lymphocytes) in which TANGO 358 is expressed. TANGO 358 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

Human TANGO 365

A cDNA encoding TANGO 365 was identified by analyzing the sequences of clones present in a human prostate fibroblast library for sequences that encode wholly secreted or transmembrane proteins. This analysis led to the identification of a clone, jthqc001g06, encoding full-length Human TANGO 365. The TANGO 365 cDNA of this clone is 1338 nucleotides long (Figure 9; SEQ ID NO:44). The open reading frame of this cDNA, nucleotides 56 to 550 of SEQ ID NO:44 (SEQ ID NO:45), encodes a 165 amino acid transmembrane protein (Figure 9; SEQ ID NO:46).

Figure 10 depicts a hydropathy plot of human TANGO 365. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively

hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 36 of SEQ ID NO:46; SEQ ID NO:47) on the left from the mature protein (amino acids 37 to 165 of SEQ ID NO:46; SEQ ID NO:48) on the right.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 365 includes a 36 amino acid signal peptide (amino acid 1 to amino acid 36 of SEQ ID NO:46; SEQ ID NO:47) preceding the mature protein (corresponding to amino acid 37 to amino acid 165 of SEQ ID NO:46; SEQ ID NO:48). The molecular weight of TANGO 365 protein without post-translational modifications is 17.4 kDa prior to the cleavage of the signal peptide, 13.6 kDa after cleavage of the signal peptide. The presence of a methionine residue at positions 16, 35 and 81 of SEQ ID NO:46 indicates that there can be alternative forms of human TANGO 365 of 150 amino acids of SEQ ID NO:46, 131 amino acids of SEQ ID NO:46, and 65 amino acids of SEQ ID NO:46, respectively.

Human TANGO 365 is a transmembrane protein which can include one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain. The human TANGO 365 protein contains two extracellular domains; one at amino acid residues 37 to 51 of SEQ ID NO:46 (SEQ ID NO:232); and a second at amino acid residues 95 to 165 of SEQ ID NO:46 (SEQ ID NO:51), two hydrophobic transmembrane domains; one at amino acids 52 to 70 of SEQ ID NO:46 (SEQ ID NO:49); and a second at amino acids 78 to 94 of SEQ ID NO:46 (SEQ ID NO:50), and a cytoplasmic domain at amino acid residues 71 to 77 of SEQ ID NO:46 (SEQ ID NO:234).

Alternatively, in another embodiment, a human TANGO 365 protein contains two cytoplasmic domains; one at amino acid residues 37 to 51 of SEQ ID NO:46; and a second at amino acid residues 95 to 165 of SEQ ID NO:46, two hydrophobic transmembrane domains; one at amino acids 52 to 70 of SEQ ID NO:46 (SEQ ID NO:49); and a second at amino acids 78 to 94 of SEQ ID NO:46 (SEQ ID NO:50), and an extracellular domain at amino acid residues 71 to 77 of SEQ ID NO:46.

In one embodiment of a nucleotide sequence of human TANGO 365, the nucleotide at position 14 is cytosine (C)(SEQ ID NO:45). In this embodiment, the amino acid at position 5 is alanine (A)(SEQ ID NO:46). In an alternative embodiment, a species variant of human TANGO 365 has a nucleotide at position 14 which is thymidine (T)(SEQ ID NO:160). In this embodiment, the amino acid at position 5 is valine (V)(SEQ ID NO:161), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 365, the nucleotide at position 41 is guanine (G)(SEQ ID NO:45). In this embodiment, the amino acid at position 14 is arginine (R)(SEQ ID NO:46). In an alternative embodiment, a species variant of human TANGO 365 has a nucleotide at position 41 which is adenine (A)(SEQ ID NO:162). In this embodiment, the amino acid at position 14 is histidine (H)(SEQ ID NO:163), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 365, the nucleotide at position 59 is cytosine (C)(SEQ ID NO:45). In this embodiment, the amino acid at position 20 is threonine (T)(SEQ ID NO:46). In an alternative embodiment, a species variant of human TANGO 365 has a nucleotide at position 59 which is guanine (G)(SEQ ID NO:164). In this embodiment, the amino acid at position 20 is serine (S)(SEQ ID NO:165), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 365, the nucleotide at position 115 is adenine (A)(SEQ ID NO:45). In this embodiment, the amino acid at position 39 is asparagine (N)(SEQ ID NO:46). In an alternative embodiment, a species variant of human TANGO 365 has a nucleotide at position 115 which is guanine (G)(SEQ ID NO:166). In this embodiment, the amino acid at position 39 is aspartate (D)(SEQ ID NO:167), *i.e.*, a conservative substitution.

One protein kinase C phosphorylation site is present in human TANGO 365. The site has the sequence SLR and is found (at amino acids 139 to 141 of SEQ ID NO:46). The TANGO 365 protein has four N-myristoylation sites. The first has the sequence GGTRCR and is found (at amino acids 18 to 23 of SEQ ID NO:46), the second has the sequence GTSMAC and is found (at amino acids 32 to 37 of SEQ ID NO:46), the third has the sequence GAACSL and is found (at amino acids 87 to 92 of SEQ ID NO:46), and the fourth has the sequence GSSDSS and is found (at amino acids 144 to 149 of SEQ ID NO:46). Human TANGO 365 also has an amidation site which has the sequence of LGRR (at amino acids 69 to 72 of SEQ ID NO:46).

Clone EpT365, which encodes human TANGO 365, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-291. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Uses of TANGO 365 Nucleic acids, Polypeptides, and Modulators Thereof

TANGO 365 was identified as being expressed in a prostate fibroblast library. In light of this, TANGO 365 nucleic acids, proteins and modulators thereof can be utilized to diagnose disorders and/or modulate processes involved in prostate development, differentiation and activity, including, but not limited to development, and differentiation and activation of prostate tissues and cells as well as any function associated with such cells, and amelioration of one or more symptoms associated with abnormal function of such cell types. Such disorders can include, but are not limited to, malignant or benign prostate cell growth. Such disorders can include, but are not limited to, malignant or benign prostate cell growth. The TANGO 365 proteins can be used to treat subjects with or without prostate cancer *e.g.*, prostatitis, benign prostatic hypertrophy, benign prostatic hyperplasia (BPH), prostatic paraganglioma, prostate adenocarcinoma, prostatic intraepithelial neoplasia, prostato-rectal fistulas, atypical prostatic stromal lesions.

TANGO 365 nucleic acids, proteins, and modulators thereof can also be used to treat disorders of the cells and tissues in which it is expressed. As TANGO 365 is a transmembrane protein, proteins, nucleic acids and modulators thereof can be used to diagnose disorders and/or modulate intercellular signaling processes by disrupting or enhancing ligand-receptor or cell interaction with the extracellular matrix. Further, TANGO 365 could be used in detection and diagnostic assays to assay for normal or inappropriate expression of TANGO 365 proteins in aberrantly growing cells.

TANGO 365 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the prostate) and/or cells (*e.g.*, fibroblasts) in which TANGO 365 is expressed. TANGO 365 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

Human TANGO 368

A cDNA encoding human TANGO 368 was identified by analyzing the sequences of clones present in a natural killer cell library for sequences that encode wholly secreted or transmembrane proteins. This analysis led to the identification of a clone, jthta080f06, encoding full-length human TANGO 368. The human TANGO 368 cDNA of this clone is 983 nucleotides long (Figure 11; SEQ ID NO:52). The open reading frame of this cDNA, nucleotides 152 to 328 of SEQ ID NO:52 (SEQ ID NO:53), encodes a 59 amino acid secreted protein (Figure 11; SEQ ID NO:54).

Figure 12 depicts a hydropathy plot of human TANGO 368. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues

(cys) and N-glycosylation sites (NGly) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence on the left from the mature protein on the right.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 368 includes a 26 amino acid signal peptide (amino acid 1 to amino acid 27 of SEQ ID NO:54; SEQ ID NO:56) preceding the mature human TANGO 368 protein (corresponding to amino acid 28 to amino acid 59 of SEQ ID NO:54; SEQ ID NO:55). The molecular weight of TANGO 368 protein without post-translational modifications is 6.5 kDa prior to the cleavage of the signal peptide and 3.5 kDa after cleavage of the signal peptide.

In one embodiment of a nucleotide sequence of human TANGO 368, the nucleotide at position 8 is cytosine (C)(SEQ ID NO:53). In this embodiment, the amino acid at position 3 is threonine (T)(SEQ ID NO:54). In an alternative embodiment, a species variant of human TANGO 368 has a nucleotide at position 8 which is guanine (G)(SEQ ID NO:168). In this embodiment, the amino acid at position 3 is serine (S)(SEQ ID NO:169), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 368, the nucleotide at position 10 is cytosine (C)(SEQ ID NO:53). In this embodiment, the amino acid at position 4 is glutamine (Q)(SEQ ID NO:54). In an alternative embodiment, a species variant of human TANGO 368 has a nucleotide at position 10 which is guanine (G)(SEQ ID NO:170). In this embodiment, the amino acid at position 4 is glutamate (E)(SEQ ID NO:171), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 368, the nucleotide at position 16 is cytosine (C)(SEQ ID NO:53). In this embodiment, the amino acid at position 6 is leucine (L)(SEQ ID NO:54). In an alternative embodiment, a species variant of human TANGO 368 has a nucleotide at position 16 which is guanine (G)(SEQ ID NO:172). In this embodiment, the amino acid at position 6 is valine (V)(SEQ ID NO:173), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 368, the nucleotide at position 110 is adenine (A)(SEQ ID NO:53). In this embodiment, the amino acid at position 37 is histidine (H)(SEQ ID NO:54). In an alternative embodiment, a species variant of human TANGO 368 has a nucleotide at position 110 which is guanine (G)(SEQ ID NO:174). In this embodiment, the amino acid at position 37 is arginine (R)(SEQ ID NO:175), *i.e.*, a conservative substitution.

Human TANGO 368 has an N-glycosylation site with the sequence NFTC (at amino acid residues 40 to 43 of SEQ ID NO:54), a protein kinase C phosphorylation site

with the sequence SLK (at amino acid residues 24 to 26 of SEQ ID NO:54), and a casein kinase II phosphorylation site with the sequence TQPE (at amino acid residues 27 to 30 of SEQ ID NO:54).

Figure 13 depicts a local alignment of the nucleotide sequence of full length human TANGO 368 (SEQ ID NO:52) and a fragment of the human T-cell receptor gamma V1 gene region (Accession Number AF057177; SEQ ID NO:57), which maps to a region of human chromosome 7. The full-length nucleic acid sequence of human TANGO 368 (SEQ ID NO:52) has 99.3% identity to a 973 bp fragment of the human T-cell receptor gamma V1 gene region (Accession Number AF057177; SEQ ID NO:57).

Northern blots were performed to analyze the expression of human TANGO 368 mRNA in human tissues. A weak signal was observed in the spleen and lymph node, however, no expression was detected in the thymus, peripheral blood leukocytes or fetal liver.

Clone EpT368, which encodes human TANGO 368, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-291. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Uses of TANGO 368 Nucleic acids, Polypeptides, and Modulators Thereof

As TANGO 368 was originally found in a natural killer cell library, TANGO 368 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, development, differentiation, and/or function of immune cells, such as lymphocytes, *e.g.*, natural killer cells, T-cells and B-cells. TANGO 368 nucleic acids, proteins and modulators thereof can be utilized to modulate immune-related processes *e.g.*, the host immune response by, for example, modulating the formation of and/or binding to immune complexes, detection and defense against surface antigens and bacteria, and immune surveillance for rapid removal or pathogens. Such TANGO 368 nucleic acids, proteins and modulators thereof can be utilized, *e.g.*, to ameliorate incidence of any symptoms associated with disorders that involve such immune-related processes, including, but not limited to viral or bacterial infection, autoimmune disorders (*e.g.*, Grave's disease, Hashimoto's disease, and arthritis), immunodeficiency disorders (*e.g.*, HIV, and inflammatory disorders (*e.g.*, asthma, arthritis, psoriasis, septicemia, inflammatory bowel disease and allergies).

As TANGO 368 exhibits expression in the spleen, TANGO 368 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, differentiation, and/or function of cells that form the spleen, *e.g.*, cells of the splenic connective tissue, *e.g.*, splenic smooth muscle cells and/or endothelial cells of the splenic blood vessels. TANGO 368 nucleic acids, proteins, and modulators thereof can also be used to modulate the proliferation, differentiation, and/or function of cells that are processed, *e.g.*, regenerated or phagocytized within the spleen, *e.g.*, erythrocytes and/or B and T lymphocytes and macrophages. Thus, TANGO 368 nucleic acids, proteins, and modulators thereof can be used to treat spleen, *e.g.*, the fetal spleen, associated diseases and disorders. Examples of splenic diseases and disorders include *e.g.*, splenic lymphoma and/or splenomegaly, and/or phagocytotic disorders, *e.g.*, those inhibiting macrophage engulfment of bacteria and viruses in the bloodstream.

As TANGO 368 exhibits expression in the lymph nodes, TANGO 368 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, differentiation, and/or function of cells that form the lymph node, *e.g.*, cells of the lymph node connective tissue, *e.g.*, lymph node smooth muscle cells and/or endothelial cells of the lymph node blood vessels. TANGO 368 nucleic acids, proteins, and modulators thereof can also be used to diagnose disorders and/or modulate the proliferation, differentiation, and/or function of cells that are processed, *e.g.*, phagocytized within the lymph node, *e.g.*, erythrocytes and/or B and T lymphocytes and macrophages. Thus, TANGO 368 nucleic acids, proteins, and modulators thereof can be used to treat lymph node associated diseases and disorders. Examples of lymph node diseases and disorders include *e.g.*, lymphadenopathy, lymphoma, and/or phagocytotic disorders, *e.g.*, those inhibiting macrophage engulfment of bacteria and viruses in the bloodstream.

In light of the fact that TANGO 368 is homologous to the T-cell receptor gamma (TCRγ) locus, TANGO 368 nucleic acids, proteins and modulators thereof can be utilized to modulate the recognition of antigens in association with the major histocompatibility complex TANGO 368 nucleic acids, proteins and modulators thereof can be utilized to modulate diseases and/or disorders associated with aberrant TCR-MHC interactions. Further, TANGO 368 nucleic acids, proteins and modulators thereof can be utilized to modulate cell-cell receptor interactions.

As TANGO 368 exhibits homology to human T-cell receptor gamma V1 gene region (Accession Numbers AF057177), which maps to a region of chromosome 7, TANGO 368 nucleic acids, proteins and modulators thereof can be utilized to diagnose disorders and/or modulate diseases associated with that region of chromosome 7, *e.g.*, Stiff-Mann syndrome.

As TANGO 368 is a secreted protein and thus likely a signaling molecule, TANGO 368 nucleic acids, proteins or modulators thereof, can be used to modulate TANGO 368 biological activities, which include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) ability to modulate cell-cell interactions; (3) the ability to modulate the proliferation, differentiation and/or activity of neural cells; and (4) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades).

TANGO 368 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the thymus) and/or cells (*e.g.*, natural killer cells) in which TANGO 368 is expressed. TANGO 368 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

Human TANGO 369

A cDNA encoding human TANGO 369 was identified by analyzing the sequences of clones present in a natural killer cell library for sequences that encode wholly secreted or transmembrane proteins. This analysis led to the identification of a clone, jthta088h08, encoding full-length human TANGO 369. The human TANGO 369 cDNA of this clone is 1119 nucleotides long (Figure 14; SEQ ID NO:58). The open reading frame of this cDNA, nucleotides 162 to 335 of SEQ ID NO:58 (SEQ ID NO:59), encodes a 58 amino acid secreted protein (Figure 14; SEQ ID NO:60).

Figure 15 depicts a hydropathy plot of human TANGO 369. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence on the left from the mature protein on the right.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 369 includes a 26 amino acid signal peptide (amino acid 1 to amino acid 26 of SEQ ID NO:60; SEQ ID NO:62) preceding the mature human TANGO 369 protein (corresponding to amino acid 27 to amino acid 58 of SEQ ID NO:60; SEQ ID NO:61). The molecular weight of TANGO 369 protein without post-translational modifications is 6.8 kDa prior to the cleavage of the signal peptide and 3.7 kDa after cleavage of the signal peptide. The presence of a methionine residue at positions 17 and 25 of SEQ ID NO:60 indicates that there can be alternative forms of human TANGO 369 of 42 amino acids of SEQ ID NO:60, and 34 amino acids of SEQ ID NO:60, respectively.

In one embodiment of a nucleotide sequence of human TANGO 369, the nucleotide at position 58 is cytosine (C)(SEQ ID NO:59). In this embodiment, the amino acid at position 20 is leucine (L)(SEQ ID NO:54). In an alternative embodiment, a species variant of human TANGO 369 has a nucleotide at position 58 which is guanine (G)(SEQ ID NO:176). In this embodiment, the amino acid at position 20 is valine (V)(SEQ ID NO:177), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 369, the nucleotide at position 68 is guanine (G)(SEQ ID NO:59). In this embodiment, the amino acid at position 23 is arginine (R)(SEQ ID NO:60). In an alternative embodiment, a species variant of human TANGO 369 has a nucleotide at position 68 which is adenine (A)(SEQ ID NO:178). In this embodiment, the amino acid at position 23 is lysine (K)(SEQ ID NO:179), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 369, the nucleotide at position 70 is thymine (T)(SEQ ID NO:59). In this embodiment, the amino acid at position 24 is leucine (L)(SEQ ID NO:60). In an alternative embodiment, a species variant of human TANGO 369 has a nucleotide at position 70 which is adenine (A)(SEQ ID NO:180). In this embodiment, the amino acid at position 24 is threonine (T)(SEQ ID NO:181), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 369, the nucleotide at position 120 is guanine (G)(SEQ ID NO:59). In this embodiment, the amino acid at position 40 is glutamate (E)(SEQ ID NO:60). In an alternative embodiment, a species variant of human TANGO 369 has a nucleotide at position 120 which is cytosine (C)(SEQ ID NO:182). In this embodiment, the amino acid at position 40 is aspartate (D)(SEQ ID NO:183), *i.e.*, a conservative substitution.

Northern blots were performed to analyze the expression of human TANGO 369 mRNA in human tissues. A very weak signal was observed in the spleen and lymph node, however, no expression was detected in the thymus, peripheral blood leukocytes or fetal liver.

Clone EpT369, which encodes human TANGO 369, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-295. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Uses of TANGO 369 Nucleic acids, Polypeptides, and Modulators Thereof

As TANGO 369 was originally found in a natural killer cell library, TANGO 369 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, development, differentiation, and/or function of lymphocytes, *e.g.*, natural killer cells. TANGO 369 nucleic acids, proteins and modulators thereof can be utilized to modulate immune-related processes, *e.g.*, the host immune response by, for example, modulating the formation of and/or binding to immune complexes, detection and defense against surface antigens and bacteria, and immune surveillance for rapid removal or pathogens. Such TANGO 369 compositions and modulators thereof can be utilized, *e.g.*, to ameliorate incidence of any symptoms associated with disorders that involve such immune-related processes, including, but not limited to viral or bacterial infection, autoimmune disorders (*e.g.*, Grave's disease, Hashimoto's disease, arthritis, graft rejection), and inflammatory disorders (*e.g.*, bacterial or viral infection, psoriasis, allergies and inflammatory bowel diseases).

As TANGO 369 exhibits expression in the spleen, TANGO 369 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, differentiation, and/or function of cells that form the spleen, *e.g.*, cells of the splenic connective tissue, *e.g.*, splenic smooth muscle cells and/or endothelial cells of the splenic blood vessels. TANGO 369 nucleic acids, proteins, and modulators thereof can also be used to modulate the proliferation, differentiation, and/or function of cells that are processed, *e.g.*, regenerated or phagocytized within the spleen, *e.g.*, erythrocytes and/or B and T lymphocytes and macrophages. Thus, TANGO 369 nucleic acids, proteins, and modulators thereof can be used to treat spleen, *e.g.*, the fetal spleen, associated diseases and disorders. Examples of splenic diseases and disorders include *e.g.*, splenic lymphoma and/or splenomegaly, and/or phagocytotic disorders, *e.g.*, those inhibiting macrophage engulfment of bacteria and viruses in the bloodstream.

As TANGO 369 exhibits expression in the lymph nodes, TANGO 369 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, differentiation, and/or function of cells that form the lymph node, *e.g.*, cells of the lymph node connective tissue, *e.g.*, lymph node smooth muscle cells and/or endothelial cells of the lymph node blood vessels. TANGO 369 nucleic acids, proteins, and modulators thereof can also be used to modulate the proliferation, differentiation, and/or function of cells that are processed, *e.g.*, phagocytized within the lymph node, *e.g.*, erythrocytes and/or B and T lymphocytes and macrophages. Thus, TANGO 369 nucleic acids, proteins, and modulators thereof can be used to treat lymph node associated diseases and disorders. Examples of lymph node diseases and disorders include *e.g.*,

lymphadenopathy, lymphoma, and/or phagocytotic disorders, *e.g.*, those inhibiting macrophage engulfment of bacteria and viruses in the bloodstream.

TANGO 369 is associated with immune cells. As such, immune disorders associated TANGO 369 nucleic acids, proteins and modulators thereof can be used to diagnose disorders and/or modulate or treat immune disorders that include, but are not limited to, immune proliferative disorders (*e.g.*, carcinoma, lymphoma, *e.g.*, follicular lymphoma), and disorders associated with fighting pathogenic infections, *e.g.*, bacterial (*e.g.*, chlamydia) infection, parasitic infection, and viral infection (*e.g.*, HSV infection), and pathogenic disorders associated with immune disorders (*e.g.*, immunodeficiency disorders, such as HIV).

Other immune disorders associated with TANGO 369 activity, for which TANGO 369 nucleic acids, proteins and modulators thereof can be used to modulate, identify, diagnose or treat, include, *e.g.*, autoimmune disorders, such as arthritis, graft rejection (*e.g.*, allograft rejection), T cell disorders (*e.g.*, AIDS)) and inflammatory disorders, such as bacterial infection, psoriasis, septicemia, cerebral malaria, inflammatory bowel disease, arthritis (*e.g.*, rheumatoid arthritis, osteoarthritis), and allergic inflammatory disorders (*e.g.*, asthma, psoriasis), apoptotic disorders (*e.g.*, rheumatoid arthritis, systemic lupus erythematosus, insulin-dependent diabetes mellitus), cytotoxic disorders, septic shock, cachexia, and proliferative disorders (*e.g.*, B cell cancers stimulated by TNF).

Other TANGO 369 associated immune disorders include TNF related disorders (*e.g.*, acute myocarditis, myocardial infarction, congestive heart failure, T cell disorders (*e.g.*, dermatitis, fibrosis)), differentiative and apoptotic disorders, and disorders related to angiogenesis (*e.g.*, tumor formation and/or metastasis, cancer). TANGO 369 nucleic acids, proteins and modulators thereof can be used to treat such disorders.

As TANGO 369 is a secreted protein, TANGO 369 nucleic acids, proteins and modulators thereof can be utilized to modulate intercellular signaling pathways, for example by disrupting ligand-receptor interactions or cellular interactions with the extra-cellular matrix.

As TANGO 369 is a secreted protein and thus likely a signaling molecule, TANGO 369 nucleic acids, proteins or modulators thereof can be used TANGO 369 biological activities, which can also include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) ability to modulate cell-cell interactions; (3) the ability to modulate the proliferation, differentiation and/or activity of neural cells; and (4) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades).

TANGO 369 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the thymus) and/or cells (*e.g.*, natural killer cells) in which TANGO 369 is expressed. TANGO 369 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

Human TANGO 383

A cDNA encoding human TANGO 383 was identified by analyzing the sequences of clones present in a human prostate epithelium cDNA library. This analysis led to the identification of a clone, jthqb083b10, encoding full-length TANGO 383. The human cDNA of this clone is 1386 nucleotides long (Figure 16; SEQ ID NO:63). The open reading frame of this cDNA, nucleotides 104 to 523 of SEQ ID NO:63 (SEQ ID NO:64), encodes a 140 amino acid TANGO 383 transmembrane protein (Figure 16; SEQ ID NO:65).

Figure 17 depicts a hydropathy plot of human TANGO 383. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 20 of SEQ ID NO:65; SEQ ID NO:66) on the left from the mature protein (amino acids 21 to 140 of SEQ ID NO:65; SEQ ID NO:67) on the right.

The signal peptide prediction program SIGNALP (Nielsen, et al. (1997) *Protein Engineering* 10:1-6) predicted that TANGO 383 includes a 20 amino acid signal peptide (amino acid 1 to amino acid 20 of SEQ ID NO:65; SEQ ID NO:66) preceding the mature protein (corresponding to amino acid 21 to amino acid 140 of SEQ ID NO:65; SEQ ID NO:67). The molecular weight of TANGO 383 without post-translational modifications is 14.9 kDa prior to the cleavage of the signal peptide, 12.7 kDa after cleavage of the signal peptide.

TANGO 383 is a transmembrane protein which contains one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain. The TANGO 383 protein contains an extracellular domain at amino acids 71 to 115 of SEQ ID NO:65 (SEQ ID NO:70), a first transmembrane domain at amino acid residues 50 to 70 of SEQ ID NO:65 (SEQ ID NO:68), a second transmembrane domain at amino acid residues 116 to 133 of SEQ ID NO:65 (SEQ ID NO:69), a first cytoplasmic domain at amino acid residues 21 to 49 of SEQ ID NO:65 (SEQ ID NO:235) and a second cytoplasmic domain at amino acid residues 134 to 140 of SEQ ID NO:65 (SEQ ID NO:136).

Alternatively, in another embodiment, a TANGO 383 protein contains a cytoplasmic domain at amino acids 71 to 115 of SEQ ID NO:65, a first transmembrane domain at amino acid residues 50 to 70 of SEQ ID NO:65 (SEQ ID NO:68), a second transmembrane domain at amino acid residues 116 to 133 of SEQ ID NO:65 (SEQ ID NO:69), a first extracellular domain at amino acid residues 21 to 49 of SEQ ID NO:65 and a second extracellular domain at amino acid residues 134 to 140 of SEQ ID NO:65.

In one embodiment of a nucleotide sequence of human TANGO 383, the nucleotide at position 4 is cytosine (C)(SEQ ID NO:64). In this embodiment, the amino acid at position 2 is leucine (L)(SEQ ID NO:65). In an alternative embodiment, a species variant of human TANGO 383 has a nucleotide at position 4 which is adenine (A)(SEQ ID NO:184). In this embodiment, the amino acid at position 2 is isoleucine (I)(SEQ ID NO:185), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 383, the nucleotide at position 8 is guanine (G)(SEQ ID NO:64). In this embodiment, the amino acid at position 3 is serine (S)(SEQ ID NO:65). In an alternative embodiment, a species variant of human TANGO 383 has a nucleotide at position 8 which is cytosine (C)(SEQ ID NO:186). In this embodiment, the amino acid at position 3 is threonine (T)(SEQ ID NO:187), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 383, the nucleotide at position 17 is adenine (A)(SEQ ID NO:64). In this embodiment, the amino acid at position 6 is lysine (K)(SEQ ID NO:65). In an alternative embodiment, a species variant of human TANGO 383 has a nucleotide at position 17 which is guanine (G)(SEQ ID NO:188). In this embodiment, the amino acid at position 6 is arginine (R)(SEQ ID NO:189), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 383, the nucleotide at position 57 is cytosine (C)(SEQ ID NO:64). In this embodiment, the amino acid at position 19 is aspartate (D)(SEQ ID NO:65). In an alternative embodiment, a species variant of human TANGO 383 has a nucleotide at position 57 which is guanine (G)(SEQ ID NO:190). In this embodiment, the amino acid at position 19 is glutamate (E)(SEQ ID NO:191), *i.e.*, a conservative substitution.

One protein kinase C phosphorylation site is present in TANGO 383, and has the sequence SPR (at amino acids 21 to 24 of SEQ ID NO:65). TANGO 383 has one casein kinase II phosphorylation site which has the sequence SKAE (at amino acids 42 to 45 of SEQ ID NO:65). TANGO 383 has three N-myristylation sites. The first has the sequence GVELAS (at amino acids 24 to 29 of SEQ ID NO:65), the second has the sequence GAVLAH (at amino acids 84 to 89 of SEQ ID NO:65), and the third has the sequence

GSSDSH (at amino acids 96 to 101 of SEQ ID NO:65). TANGO 383 has a consensus tyrosine phosphorylation site which has the amino acid sequence RGKREAGLY and (at amino acids 33 to 41 of SEQ ID NO:65). TANGO 383 also has an amidation site with the sequence RGKR (at amino acids 33-36 of SEQ ID NO:65).

Figure 18 depicts an alignment of the amino acid sequence of TANGO 383 (SEQ ID NO:65) and the amino acid sequence of Neuronal Thread Protein AD7C-NTP (SEQ ID NO:72). The alignments demonstrates that the amino acid sequences of TANGO 383 and Neuronal Thread Protein AD7C-NTP are 52% identical. This alignment was performed using the ProDom NCBI-BLASTP2 program with graphical output using the following settings: Matrix: BLOSUM62; Expect: 0.1; Filter: none.

Thus, TANGO 383 exhibits homology to neural thread proteins which are phospho-proteins expressed in the central nervous system which are phosphorylated during neuritic sprouting. Therefore, TANGO 383 nucleic acids, proteins and modulators thereof may be used to diagnose disorders and/or inhibit or modulate neurodegenerative sprouting and synaptic disassociation associated with, *e.g.*, Alzheimer's disease, and other diseases in neural tissue as discussed below.

Clone EpT383, which encodes human TANGO 383, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-295. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Uses of TANGO 383 Nucleic acids, Polypeptides, and Modulators Thereof

As TANGO 383 was originally found in a prostate epithelium library, TANGO 383 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, differentiation, and/or function of prostate cells. TANGO 383 nucleic acids, proteins and modulators thereof can be utilized to modulate processes involved in prostate development, differentiation and activity, including, but not limited to development, and differentiation and activation of prostate tissues and cells as well as any function associated with such cells, and amelioration of one or more symptoms associated with abnormal function of such cell types or disorders associated with such cell types. Such disorders can include, but are not limited to, malignant or benign prostate cell growth or inflammatory disorders (*e.g.*, prostatitis, benign prostatic hypertrophy, benign prostatic hyperplasia (BPH), prostatic paraganglioma, prostate

adenocarcinoma, prostatic intraepithelial neoplasia, prostato-rectal fistulas, atypical prostatic stromal lesions).

TANGO 383 exhibits homology to neural thread proteins which are phosphoproteins expressed in the central nervous system which are phosphorylated during neuritic sprouting. Therefore, TANGO 383 nucleic acids, proteins and modulators thereof may be used to diagnose disorders and/or inhibit or modulate neurodegenerative sprouting and synaptic disassociation associated with, *e.g.*, Alzheimer's disease. TANGO 383 nucleic acids, proteins and modulators thereof may also be utilized to diminish the effects of stroke and other neural damage, *e.g.*, spinal cord injuries, infarction, infection, malignancy, exposure to toxic agents, nutritional deficiency, paraneoplastic syndromes, and degenerative nerve diseases including but not limited to Alzheimer's disease, Parkinson's disease, Huntington's Chorea, amyotrophic lateral sclerosis, progressive supranuclear palsy, and other dementias.

As TANGO 383 is a transmembrane protein, TANGO 383 nucleic acids, proteins and modulators thereof can be utilized to modulate intercellular signaling pathways, for example by disrupting ligand-receptor interactions or cellular interactions with the extracellular matrix.

As TANGO 383 is a transmembrane protein and thus likely a signaling molecule, TANGO 383 nucleic acids, proteins or modulators thereof, activities can include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) ability to modulate cell-cell interactions; (3) the ability to modulate the proliferation, differentiation and/or activity of neural cells; and (4) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades).

TANGO 383 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the prostate) and/or cells (*e.g.*, epithelial cells) in which TANGO 383 is expressed. TANGO 383 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

<u>Human TANGO 393</u>

A cDNA encoding human TANGO 393 was identified by analyzing the sequences of clones present in a human fetal hypothalamus cDNA library for sequences containing signal peptides. This analysis led to the identification of a clone, jthhb039f09, encoding full-length human TANGO 393. The human cDNA of this clone is 1778 nucleotides long (Figure 19; SEQ ID NO:73). The open reading frame of this cDNA, nucleotides 40 to

1458 of SEQ ID NO:75 (SEQ ID NO:74), encodes a 473 amino acid human TANGO 393 transmembrane protein (Figure 19; SEQ ID NO:75).

Figure 20 depicts a hydropathy plot of human TANGO 393. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) and potential N-glycosylation sites are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 26 of SEQ ID NO:75; SEQ ID NO:76) on the left from the mature protein (amino acids 27 to 473 of SEQ ID NO:75; SEQ ID NO:77) on the right.

The signal peptide prediction program SIGNALP (Nielsen, et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 393 includes an 26 amino acid signal peptide (amino acid 1 to amino acid 26 of SEQ ID NO:75; SEQ ID NO:76) preceding the mature protein (corresponding to amino acid 27 to amino acid 473 of SEQ ID NO:75; SEQ ID NO:77). The molecular weight of human TANGO 393 without post-translational modifications is 50.7 kDa prior to the cleavage of the signal peptide, 47.8 kDa after cleavage of the signal peptide. The presence of a methionine residue at position 229 of SEQ ID NO:75 indicates that there can be alternative forms of human TANGO 393 of 245 amino acids of SEQ ID NO:75.

Human TANGO 393 is a transmembrane protein which contains one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain; and (4) a leucine-rich domain. The human TANGO 393 protein contains an extracellular domain at amino acids 27 to 447 of SEQ ID NO:75 (SEQ ID NO:89), a transmembrane domain at amino acid residues 448 to 467 of SEQ ID NO:75 (SEQ ID NO:78), and a cytoplasmic domain at amino acid residues 468 to 473 of SEQ ID NO:75 (SEQ ID NO:134).

Alternatively, in another embodiment, a human TANGO 393 protein contains a cytoplasmic domain at amino acids 27 to 447 of SEQ ID NO:75, a transmembrane domain at amino acid residues 448 to 467 of SEQ ID NO:75 (SEQ ID NO:78), and a extracellular domain at amino acid residues 468 to 473 of SEQ ID NO:75.

In one embodiment of a nucleotide sequence of human TANGO 393, the nucleotide at position 5 is adenine (A)(SEQ ID NO:74). In this embodiment, the amino acid at position 2 is lysine (K)(SEQ ID NO:75). In an alternative embodiment, a species variant of human TANGO 393 has a nucleotide at position 5 which is guanine (G)(SEQ ID NO:192). In this embodiment, the amino acid at position 2 is arginine (R)(SEQ ID NO:193), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 393, the nucleotide at position 17 is cytosine (C)(SEQ ID NO:74). In this embodiment, the amino acid at position 6 is alanine (A)(SEQ ID NO:75). In an alternative embodiment, a species variant of human TANGO 393 has a nucleotide at position 17 which is thymidine (T)(SEQ ID NO:194). In this embodiment, the amino acid at position 6 is valine (V)(SEQ ID NO:195), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 393, the nucleotide at position 55 is cytosine (C)(SEQ ID NO:74). In this embodiment, the amino acid at position 19 is glutamine (Q)(SEQ ID NO:75). In an alternative embodiment, a species variant of human TANGO 393 has a nucleotide at position 55 which is guanine (G)(SEQ ID NO:196). In this embodiment, the amino acid at position 19 is glutamate (E)(SEQ ID NO:197), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 393, the nucleotide at position 118 is adenine (A)(SEQ ID NO:74). In this embodiment, the amino acid at position 40 is threonine (T)(SEQ ID NO:75). In an alternative embodiment, a species variant of human TANGO 393 has a nucleotide at position 118 which is thymine (T)(SEQ ID NO:198). In this embodiment, the amino acid at position 40 is serine (S)(SEQ ID NO:199), *i.e.*, a conservative substitution.

Human TANGO 393 has LRR from amino acids 26 to 57, 58 to 81, 82 to 105, 106 to 130, 131 to 154, 155 to 178, 179 to 202, 203 to 226, 227 to 250, and 260 to 310 of SEQ ID NO:75 (SEQ ID NO:79, 80, 81, 82, 83, 84, 85, 86, 87 and 88, respectively). These repeats are spaced in beta-alpha folds in the structure of the protein, so as to create a hydrophobic face that induces particular folding of the protein.

Human TANGO 393 has five N-glycosylation sites. The first has a sequence of NLTI (at amino acids 82-85 of SEQ ID NO:75), the second has a sequence of NLTH (at amino acids 179 to 182 of SEQ ID NO:75), the third has a sequence of NLSA (at amino acids 237 to 240 of SEQ ID NO:75), the fourth has a sequence of NGSG (at amino acids 372 to 375 of SEQ ID NO:75), and the fifth has a sequence of NRTR (at amino acids 423 to 426 of SEQ ID NO:75). Human TANGO 393 has one Glycosaminoglycan attachment site, the sequence of which is SGGG (at amino acids 436 to 439 of SEQ ID NO:75). Human TANGO 393 has one cAMP and cGMP-dependent protein kinase phosphorylation site, the sequence of which is KRAS (at amino acids 2 to 5 of SEQ ID NO:75). Human TANGO 393 has five protein kinase C phosphorylation sites, where the first has a sequence SQR of (at amino acids 59 to 61 of SEQ ID NO:75), the second has a sequence SFR of (at amino acids 76 to 78 of SEQ ID NO:75), the third has a sequence TFR of (at amino acids 173 to 175 of SEQ ID NO:75), the fourth has a sequence TGR of (at amino

acids 321 to 323 of SEQ ID NO:75), and the fifth has a sequence SRK of (at amino acids 420 to 422 of SEQ ID NO:75). Human TANGO 393 has five casein kinase II phosphorylation sites, where the first has a sequence of TFRD (at amino acids 173 to 176 of SEQ ID NO:75), the second has a sequence of SVPE (at amino acids 192 to 195 of SEQ ID NO:75), the third has a sequence of SSSE (at amino acids 281 to 284 of SEQ ID NO:75), the fourth has a sequence of TDEE (at amino acids 325 to 328 of SEQ ID NO:75), and the fifth has a sequence of SVLE (at amino acids 345 to 348 of SEQ ID NO:75). Human TANGO 393 has eleven N-myristylation sites, where the first has the sequence GACVCY (at amino acids 29 to 34 of SEQ ID NO:75), the second has the sequence GIPAAS (at amino acids 54 to 59 of SEQ ID NO:75), and the third has the sequence GNRISH (at amino acids 66 to 71 of SEQ ID NO:75), the fourth has the sequence GLFRGL (at amino acids 148 to 153 of SEQ ID NO:75), and the fifth has the sequence GNRISS (at amino acids 187 to 192 of SEQ ID NO:75), the sixth has the sequence GCAVAT (at amino acids 308 to 313 of SEQ ID NO:75), and the seventh has the sequence GLPKCC (at amino acids 331 to 336 of SEQ ID NO:75), the eighth has the sequence GTLPGS (at amino acids 385 to 390 of SEQ ID NO:75), and the ninth has the sequence GQAGSG (at amino acids 432 to 437 of SEQ ID NO:75), the tenth has the sequence GGGTGD (at amino acids 438 to 443 of SEQ ID NO:75), and the eleventh has the sequence GALPSL (at amino acids 448 to 453 of SEQ ID NO:75). Human TANGO 393 has a Leucine zipper pattern which has the amino acid sequence LHLDRCGLQELGPGLFRGLAAL (at amino acids 135 to 156 of SEQ ID NO:75).

Human TANGO 393 maps by homology to ESTs to Chromosome 22 between D22S420 and D22S446.

Clone EpT393, which encodes human TANGO 393, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-295. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. § 112.

<u>Mouse TANGO 393</u>

A cDNA encoding mouse TANGO 393 was identified in an analysis of a fetal hypothalamus library for screening encoding signal peptides. This analysis led to the identification of a clone, jtmoa038d08, encoding full-length mouse TANGO 393. The mouse cDNA of this clone is 1946 nucleotides long (Figure 21; SEQ ID NO:93). The

open reading frame is from nucleotides 226 to 1644 of SEQ ID NO:93 (SEQ ID NO:94), encodes a 473 amino acid mouse TANGO 393 transmembrane protein (Figure 21; SEQ ID NO:95).

Figure 22 depicts a hydropathy plot of mouse TANGO 393. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) and potential N-glycosylation sites are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 26 of SEQ ID NO:95; SEQ ID NO:96) on the left from the mature protein (amino acids 27 to 473 of SEQ ID NO:95; SEQ ID NO:97) on the right.

The signal peptide prediction program SIGNALP (Nielsen, et al. (1997) *Protein Engineering* 10:1-6) predicted that mouse TANGO 393 includes an 26 amino acid signal peptide (amino acid 1 to amino acid 26 of SEQ ID NO:95; SEQ ID NO:96) preceding the mature protein (corresponding to amino acid 27 to amino acid 473 of SEQ ID NO:95; SEQ ID NO:97). The molecular weight of mouse TANGO 393 without post-translational modifications is 51.0 kDa prior to the cleavage of the signal peptide, 48.1 kDa after cleavage of the signal peptide. The presence of a methionine residue at positions 229, 240 and 247 of SEQ ID NO:95 indicates that there can be alternative forms of mouse TANGO 393 of 245 amino acids of SEQ ID NO:95, 234 amino acids of SEQ ID NO:95, and 227 amino acids of SEQ ID NO:95, respectively.

Mouse TANGO 393 is a transmembrane protein which contains one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; (3) a cytoplasmic domain; and (4) leucine-rich repeat domain. The mouse TANGO 393 protein contains an extracellular domain at amino acids 27 to 449 of SEQ ID NO:95 (SEQ ID NO:109), a transmembrane domain at amino acid residues 450 to 467 of SEQ ID NO:75 (SEQ ID NO:98), and a cytoplasmic domain at amino acid residues 468 to 473 of SEQ ID NO:95 (SEQ ID NO:135).

Alternatively, in another embodiment, a mouse TANGO 393 protein contains a cytoplasmic domain at amino acids 27 to 449 of SEQ ID NO:95, a transmembrane domain at amino acid residues 450 to 467 of SEQ ID NO:75 (SEQ ID NO:98), and an extracellular domain at amino acid residues 468 to 473 of SEQ ID NO:95.

In one embodiment of a nucleotide sequence of mouse TANGO 393, the nucleotide at position 5 is adenine (A)(SEQ ID NO:94). In this embodiment, the amino acid at position 2 is lysine (K)(SEQ ID NO:95). In an alternative embodiment, a species variant of mouse TANGO 393 has a nucleotide at position 5 which is guanine (G)(SEQ ID

NO:200). In this embodiment, the amino acid at position 2 is arginine (R)(SEQ ID NO:201), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of mouse TANGO 393, the nucleotide at position 59 is cytosine (C)(SEQ ID NO:94). In this embodiment, the amino acid at position 20 is alanine (A)(SEQ ID NO:95). In an alternative embodiment, a species variant of mouse TANGO 393 has a nucleotide at position 59 which is thymidine (T)(SEQ ID NO:202). In this embodiment, the amino acid at position 20 is valine (V)(SEQ ID NO:203), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of mouse TANGO 393, the nucleotide at position 118 is adenine (A)(SEQ ID NO:94). In this embodiment, the amino acid at position 40 is threonine (T)(SEQ ID NO:95). In an alternative embodiment, a species variant of mouse TANGO 393 has a nucleotide at position 118 which is thymidine (T)(SEQ ID NO:204). In this embodiment, the amino acid at position 40 is serine (S)(SEQ ID NO:205), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of mouse TANGO 393, the nucleotide at position 178 is cytosine (C)(SEQ ID NO:94). In this embodiment, the amino acid at position 60 is glutamine (Q)(SEQ ID NO:95). In an alternative embodiment, a species variant of mouse TANGO 393 has a nucleotide at position 178 which is guanine (G)(SEQ ID NO:206). In this embodiment, the amino acid at position 60 is glutamate (E)(SEQ ID NO:207), *i.e.*, a conservative substitution.

Mouse TANGO 393 has five N-glycosylation sites. The first has a sequence of NLTI (at amino acids 82-85 of SEQ ID NO:95), the second has a sequence of NLTH (at amino acids 179 to 182 of SEQ ID NO:95), the third has a sequence of NLSM (at amino acids 237 to 240 of SEQ ID NO:95), the fourth has a sequence of NGSG (at amino acids 372 to 375), and the fifth has a sequence of NRTR (at amino acids 423 to 426 of SEQ ID NO:95). Mouse TANGO 393 has one Glycosaminoglycan attachment site, the sequence of which is SGTG (at amino acids 439 to 442 of SEQ ID NO:95). Mouse TANGO 393 has one cAMP- and cGMP-dependent protein kinase phosphorylation site, the sequence of which is KRAS (at amino acids 2 to 5 of SEQ ID NO:95). Mouse TANGO 393 has four protein kinase C phosphorylation sites, where the first has a sequence SQR of (at amino acids 59 to 61 of SEQ ID NO:95), the second has a sequence SCR of (at amino acids 79 to 81 of SEQ ID NO:95), the third has a sequence TFR of (at amino acids 173 to 175 of SEQ ID NO:95), and the fourth has a sequence SRK of (at amino acids 420 to 422 of SEQ ID NO:95). Mouse TANGO 393 has eight casein kinase II phosphorylation sites, where the first has a sequence of TLLE (at amino acids 105 to 108 of SEQ ID NO:95), the second has a sequence of TFRD (at amino acids 173 to 176 of SEQ ID NO:95), the third has a

sequence of SVPE (at amino acids 192 to 195 of SEQ ID NO:95), the fourth has a sequence of SSSE (at amino acids 281 to 284 of SEQ ID NO:95), the fifth has a sequence of SDLE (at amino acids 304 to 307 of SEQ ID NO:95), the sixth has a sequence of TDEE (at amino acids 325 to 328 of SEQ ID NO:95), the seventh has a sequence of SVLE (at amino acids 345 to 348 of SEQ ID NO:95), and the eighth has a sequence of SSAE (at amino acids 389 to 392 of SEQ ID NO:95). Mouse TANGO 393 has ten N-myristylation sites, where the first has the sequence GACVCY (at amino acids 29 to 34 of SEQ ID NO:95), the second has the sequence GIPAAS (at amino acids 54 to 59 of SEQ ID NO:95), and the third has the sequence GNRISH (at amino acids 66 to 71 of SEQ ID NO:95), the fourth has the sequence GLFRGL (at amino acids 148 to 153 of SEQ ID NO:95), and the fifth has the sequence GCAVAS (at amino acids 308 to 313 of SEQ ID NO:95), the sixth has the sequence GTLPSS (at amino acids 385 to 390 of SEQ ID NO:95), and the seventh has the sequence GLPTTG (at amino acids 407 to 412 of SEQ ID NO:95), the eighth has the sequence GQAGSG (at amino acids 432 to 437 of SEQ ID NO:95), and the ninth has the sequence GTGDAE (at amino acids 440 to 445 of SEQ ID NO:95), and the tenth has the sequence GALPAL (at amino acids 448 to 453 of SEQ ID NO:95). Mouse TANGO 393 has a prokaryotic membrane lipoprotein lipid attachment site with the sequence of SHVPAASFQSC (at amino acids 70 to 80 of SEQ ID NO:95). Mouse TANGO 393 has a Leucine zipper pattern which has the amino acid sequence LHLDRCGLRELGPGLFRGLAAL (at amino acids 135 to 156 of SEQ ID NO:95).

Mouse TANGO 393 has LRR from amino acids 26 to 57, 58 to 81, 82 to 105, 106 to 130, 131 to 154, 155 to 178, 179 to 202, 203 to 226, 227 to 250, and 260 to 310 of SEQ ID NO:95 (SEQ ID NO:99, 100, 101, 102, 103, 104, 105, 106, 107 and 108, respectively). These repeats are spaced in beta-alpha folds in the structure of the protein, so as to create a hydrophobic face that induces particular folding of the protein.

Figure 23 depicts an alignment of the open reading frames of human TANGO 393 (SEQ ID NO:74) and mouse TANGO 393 (SEQ ID NO:94) demonstrating an identity of 82.8%. The algorithm used to align the sequences was the ALIGN program which calculates a global alignment of two sequences. (Version 2.0u, Myers and Miller, 1989)

Figure 24 depicts an alignment of the immature proteins of human TANGO 393 (SEQ ID NO:75) and mouse TANGO 393 (SEQ ID NO:95) demonstrating an identity of 89.2%. The algorithm used to align the sequences was the ALIGN program which calculates a global alignment of two sequences. (Version 2.0u, Myers and Miller, 1989)

Uses of TANGO 393 Nucleic acids, Polypeptides, and Modulators Thereof

As both mouse and human TANGO 393 clones were originally identified in a fetal hypothalamus library, TANGO 393 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders and/or modulate the proliferation, differentiation, and/or function of endocrine cells, in particular hypothalamus, cells. TANGO 393 nucleic acids, proteins and modulators thereof can be utilized to modulate processes involved in hypothalamus development, differentiation and activity, including, but not limited to development, and differentiation and activation of hypothalamus tissues and cells as well as any function associated with such cells, and amelioration of one or more symptoms associated with abnormal function of such cell types. Such disorders can include, but are not limited to, malignant or benign hypothalamus cell growth.

Furthermore, as the hypothalamus is the master regulator of the entire endocrine system, as such, TANGO 393 nucleic acids, proteins and modulators thereof can be used as a therapeutic agent to treat mammals with abnormal hypothalamic function wherein the mammal exhibits abnormal whole animal homeostasis, appetite-related disorders, obesity, cachexia, food intake disorders, stress responsiveness disorders, adrenal function disorders, pituitary disorders and adrenal disorders. Further, TANGO 393 proteins, nucleic acids, or modulators thereof, can be used to treat disorders of the adrenal cortex, such as hypoadrenalism (*e.g.*, primary chronic or acute adrenocortical insufficiency, and secondary adrenocortical insufficiency), hyperadrenalism (Cushing's syndrome, primary hyper-aldosteronism, adrenal virilism, and adrenal hyperplasia), or neoplasia (*e.g.*, adrenal adenoma and cortical carcinoma). In another example, TANGO 393 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the thyroid gland, which is partially regulated by the hypothalamus, such as hyperthyroidism (*e.g.*, diffuse toxic hyperplasia, toxic multinodular goiter, toxic adenoma, and acute or subacute thyroiditis), hypothyroidism (*e.g.*, cretinism and myxedema), thyroiditis (*e.g.*, Hashimoto's thyroiditis, subacute granulomatous thyroiditis, subacute lymphocytic thyroiditis, Riedel's thryroiditis), Graves' disease, goiter (*e.g.*, simple diffuse goiter and multinodular goiter), or tumors (*e.g.*, adenoma, papillary carcinoma, follicular carcinoma, medullary carcinoma, undifferentiated malignant carcinoma, Hodgkin's disease, and non-Hodgkin's lymphoma).

TANGO 393 exhibits homology to genes which contain sequences referred to as Leucine Rich Repeats (LRR), for example, SLIT-1, leucine-rich α-2-Glycoprotein and Platelet Glycoprotein V precursor. As such, TANGO 393 nucleic acids, proteins and modulators thereof can be used to treat subjects with defects in leucine-rich-repeat genes shown to cause various diseases, including but not limited to Bernard-Soulier disease, a bleeding disorder. Further, as TANGO 393 has homology to Platelet Glycoprotein V (GPV) precursor, TANGO 393 nucleic acids, proteins and modulators thereof can be used

to diagnose disorders and/or modulate platelet activity, thrombin activity, von Willebrand Factor assembly and activation, or ADP/epinephrine-, cathepsin G-, and TRAP-induced decrease in platelet surface GPV expression.

Furthermore, TANGO 393 proteins, nucleic acids and modulators thereof can be used to modulate the pathogenesis of infectious diseases, for example, diseases that are affected by the expression of leucine-rich-repeat proteins such as the type-1 human immunodeficiency virus (HIV-1) Rev protein, which is the trans-activating region of the virus (Kobe and Deisenhofer, 1994, TIBS, 19:415-421).

LRR containing proteins are tissue organizers, wherein they orient and order collagen fibrils during ontogeny and are involved in pathological processes such as wound healing, tissue repair, and tumor stroma formation. These properties are rooted in their bifunctional character: the protein moiety binding collagen fibrils at strategic loci, the microscopic gaps between staggered fibrils, and the highly charged glycosaminoglycans extending out to regulate interfibrillar distances and thereby establishing the exact topology of fibrillar collagens in tissues. Therefore, TANGO 393 nucleic acids, proteins and modulators thereof can be used to disrupt intercellular and intracellular protein interactions or cellular signaling in tissues or cells, for example in the hypothalamus. More particularly, the TANGO 393 nucleic acids, proteins and modulators thereof can be used to modulate wound healing (*e.g.*, platelet activation), tissue repair and tumor stroma formation as well. Furthermore, TANGO 393 nucleic acids, proteins and modulators thereof can be used to diagnose disorders and/or modulate the function of the hypothalamus as it relates to control of endocrine function, regulation of whole animal homeostasis and modulation of diurnal requirements, appetite as related to obesity or cachexia, and generally weight control in mammals.

Proteins with LRR also interact with soluble growth factors, modulate their functional activity, and bind to cell surface receptors. The latter interaction affects cell cycle progression in a variety of cellular systems and could explain changes in the expression of these gene products around the invasive neoplastic cells and in regenerating tissues. *See Generally*, Iozzo,1997, Crit. Rev. Biochem. Mol. Biol., *32(2)*:141-74. As such, TANGO 393 nucleic acids, proteins and modulators thereof can be used to modulate disorders associated with aberrant expression of TANGO 393 in cancerous (*e.g.*, tumor) cells that do not normally express TANGO 393. Such disorders can include, for example, ones associated with tumor cell migration and progression to metastasis.

As TANGO 393 exhibits homology to the SLIT-1 proteins, TANGO 393 proteins, nucleic acids and modulators thereof may participate in the formation and maintenance of the nervous and endocrine systems by *e.g.*, protein-protein interactions. Northern blot

analysis has revealed that the human SLIT-1, -2, and -3 mRNAs are exclusively expressed in the brain, spinal cord, and thyroid, respectively. *In situ* hybridization studies indicated that the rat SLIT-1 mRNA is specifically expressed in the neurons of fetal and adult forebrains (Itoh et al., Brain Res Mol Brain Res 1998 Nov 20;62(2):175-86.) This suggests a role for TANGO 393 nucleic acids, proteins and modulators thereof in brain development and neural function. Therefore, the TANGO 393 nucleic acids, proteins and modulators thereof may be useful to disrupt protein interaction or cellular signaling in brain tissues or cells. In particular, TANGO 393 protein, nucleic acids and modulators thereof could be useful to treat neural related disorders or neural damage, such as for regenerative neural repair after damage by trauma, degeneration, or inflammation *e.g.*, multiple sclerosis, spinal cord injuries, infarction, infection, malignancy, exposure to toxic agents, nutritional deficiency, paraneoplastic syndromes, and degenerative nerve diseases including but not limited to Alzheimer's disease, Parkinson's disease, Huntington's Chorea, amyotrophic lateral sclerosis, progressive supra-nuclear palsy, and other dementia.

TANGO 393 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the hypothalamus) and/or cells (*e.g.*, hypothalamic cells) in which TANGO 393 is expressed. TANGO 393 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

Human TANGO 402

A cDNA encoding human TANGO 402 was identified by analyzing the sequences of clones present in a human 9 week fetus library for sequences that encode wholly secreted or transmembrane proteins. This analysis led to the identification of a clone, jthga055h07, encoding full-length human TANGO 402. The human TANGO 402 cDNA of this clone is 1348 nucleotides long (Figure 25; SEQ ID NO:110). The open reading frame of this cDNA, nucleotides 87 to 707 of SEQ ID NO:110 (SEQ ID NO:111), encodes a 207 amino acid transmembrane protein (Figure 25; SEQ ID NO:112).

Figure 26 depicts a hydropathy plot of human TANGO 402. Relatively hydrophobic regions of the protein are above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) and N-glycosylation sites are (Ngly) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence from the mature protein described below.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 402 includes a 50 amino acid signal peptide (amino acid 1 to amino acid 50 of SEQ ID NO:112; SEQ ID NO:114) preceding

the mature human TANGO 402 protein (corresponding to amino acid 51 to amino acid 207 of SEQ ID NO:112; SEQ ID NO:113). The molecular weight of human TANGO 402 protein without post-translational modifications is 24.0 kDa prior to the cleavage of the signal peptide, 18.1 kDa after cleavage of the signal peptide.

Human TANGO 402 protein is a transmembrane protein that contains an extracellular domain at amino acids 1 to 133 of SEQ ID NO:112 or a mature extracellular domain at amino acid residues 51 to 133 of SEQ ID NO:112 (SEQ ID NO:115), a transmembrane domain at amino acid residues 134 to 151 of SEQ ID NO:112 (SEQ ID NO:116), and a cytoplasmic domain at amino acid residues 152 to 207 of SEQ ID NO:112 (SEQ ID NO:117).

Alternatively, in another embodiment, a human TANGO 402 protein contains a cytoplasmic domain at amino acids 1 to 133 of SEQ ID NO:112 or a mature cytoplasmic domain at amino acid residues 51 to 133 of SEQ ID NO:112, a transmembrane domain at amino acid residues 134 to 151 of SEQ ID NO:112 (SEQ ID NO:116), and an extracellular domain at amino acid residues 152 to 207 of SEQ ID NO:112.

In one embodiment of a nucleotide sequence of human TANGO 402, the nucleotide at position 28 is cytosine (C)(SEQ ID NO:111). In this embodiment, the amino acid at position 10 is leucine (L)(SEQ ID NO:112). In an alternative embodiment, a species variant of human TANGO 402 has a nucleotide at position 28 which is guanine (G)(SEQ ID NO:208). In this embodiment, the amino acid at position 10 is valine (V)(SEQ ID NO:209), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 402, the nucleotide at position 58 is cytosine (C)(SEQ ID NO:111). In this embodiment, the amino acid at position 20 is glutamine (A)(SEQ ID NO:112). In an alternative embodiment, a species variant of human TANGO 402 has a nucleotide at position 58 which is guanine (G)(SEQ ID NO:210). In this embodiment, the amino acid at position 20 is glutamate (E)(SEQ ID NO:211), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 402, the nucleotide at position 61 is adenine (A)(SEQ ID NO:111). In this embodiment, the amino acid at position 21 is lysine (K)(SEQ ID NO:112). In an alternative embodiment, a species variant of human TANGO 402 has a nucleotide at position 61 which is guanine (G)(SEQ ID NO:212). In this embodiment, the amino acid at position 21 is arginine (R)(SEQ ID NO:213), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human TANGO 402, the nucleotide at position 64 is thymine (T)(SEQ ID NO:111). In this embodiment, the amino acid at position 22 is serine (S)(SEQ ID NO:112). In an alternative embodiment, a species

variant of human TANGO 402 has a nucleotide at position 64 which is adenine (A)(SEQ ID NO:214). In this embodiment, the amino acid at position 22 is threonine (T)(SEQ ID NO:215), *i.e.*, a conservative substitution.

Two N-glycosylation sites are present in human TANGO 402. The first has the sequence NISS (at amino acid residues 67 to 70 of SEQ ID NO:112) and the second has the sequence NGTS (at amino acid residues 202 to 205 of SEQ ID NO:112). Six protein kinase C phosphorylation sites are present in human TANGO 402. The first has the sequence SFK (at amino acid residues 11 to 13 of SEQ ID NO:112), the second has the sequence SFK (at amino acid residues 95 to 97 of SEQ ID NO:112), the third has the sequence TWK (at amino acid residues 98 to 100 of SEQ ID NO:112), the fourth has the sequence SQR (at amino acid residues 102 to 104 of SEQ ID NO:112), the fifth has the sequence SLK (at amino acid residues 128 to 130 of SEQ ID NO:112), and the sixth has the sequence TFK (at amino acid residues 188 to 190 of SEQ ID NO:112). Three casein kinase II phosphorylation sites are present in human TANGO 402. The first has the sequence TGID (at amino acid residues 49 to 52 of SEQ ID NO:112), the second has the sequence TWKE (at amino acid residues 98 to 101 of SEQ ID NO:112), and the third has the sequence SQRD (at amino acid residues 102 to 105 of SEQ ID NO:112). Human TANGO 402 has a tyrosine kinase phosphorylation site having the sequence KSKDFSLY at amino acid residues 21 to 28 of SEQ ID NO:112). Human TANGO 402 has an N-myristylation site having the sequence GLYVTF at amino acid residues 138 to 143 of SEQ ID NO:112.

Human TANGO 402 includes a C-type lectin (CTL)-like domain at amino acid residues 104 to 193 of SEQ ID NO:112 (SEQ ID NO:118). CTL domains have been shown to function as a calcium-dependent carbohydrate-recognition domain.

Human TANGO 402 is homologous to human lectin-like oxidized LDL receptor 1 (LOX-1), which is the receptor for oxidized lipoprotein (Sawamura et al., 1997, *Science*, 386:73-77). LOX-1 is involved in oxidized low-density lipoprotein (Ox-LDL) uptake and subsequent foam cell transformation in macrophages and smooth muscle cells in the atherosclerotic intima (Kume et al., 1998, *Cir. Res.*, 83:322-327; Yamada, et al., 1998, *Cell. Mol. Life Sci.*, 54(7):628-640; Moriwaki et al., 1998, *Artherioscler. Thromb. Vasc. Biol.*, 18(10):1541-1547; Napase et al., 1998, *J. Biol. Chem.*, 273(50):33702-33707). Figure 27 shows an alignment of the human TANGO 402 amino acid sequence (SEQ ID NO:112) with the human LOX-1 amino acid sequence (SEQ ID NO:67; Accession Number AB010710). As shown in the figure, the amino acid sequence of LOX-1 is 25.1% identical to the amino acid sequence of human TANGO 402 (SEQ ID NO:112). As shown in Figure 28. The coding regions of the human TANGO 402 nucleic acid sequence (SEQ

ID NO:56) and LOX-1 nucleic acid sequence (SEQ ID NO:66) are 42.0 % identical. The overall nucleic acid sequence identity between full-length human TANGO 402 (SEQ ID NO:110) and full-length LOX-1 (SEQ ID NO:65) is 44.0 %.

Clone EpT402, which encodes human TANGO 402, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-294. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Uses of TANGO 402 Nucleic acids, Polypeptides, and Modulators Thereof

As TANGO 402 was originally found in a human fetal library, TANGO 402 nucleic acids, proteins, and modulators thereof can be used to diagnose disorders associated with cells, tissues, and/or organs in the embryo or fetus, or modulate the proliferation, development, differentiation, and/or function of cells, tissues, and/or organs in the embryo or fetus.

In addition, as TANGO 402 is homologous to LOX-1, TANGO 402 nucleic acids, proteins and modulators thereof can be utilized to diagnose disorders, modulate development, differentiation, proliferation and/or activity of immune cells, such as macrophages and endothelial cells, *e.g.*, the phagocytosis of aged/apoptotic cells by endothelial cells. TANGO 402 nucleic acids, proteins and modulators thereof can be utilized to treat, inhibit and/or prevent disorders and diseases associated with the aberrant activity of the cells, tissues or organs in which TANGO 402 is expressed, *e.g.* endothelial activity. TANGO 402 nucleic acids, proteins and modulations thereof can also be used to diagnose disorders and/or modulate symptoms associated with atherosclerosis (*e.g.*, atherosclerotic cardiovascular disease) and Alzheimer's disease. TANGO 402 nucleic acids, proteins and modulators thereof can be used to diagnose disorders associated with host immune defenses and/or modulate host immune defenses, *e.g.*, modulating the activation of macrophages. TANGO 402 nucleic acids, proteins and modulators thereof can be utilized to treat and/or prevent obesity, diabetes, and inflammatory disorders (*e.g.*, asthma, arthritis, multiple sclerosis, allergies, hepatitis and infections).

As TANGO 402 has homology to LOX-1 proteins, TANGO 402 nucleic acids, proteins and modulators thereof can be used to modulate TANGO 402 biological activities, which include, *e.g.*, (1) the ability to bind proteins, *e.g.*, lipoproteins, *e.g.*, low-density lipoproteins, *e.g.*, oxidatively modified low-density lipoproteins; (2) the ability to

modulate internalization of proteins, *e.g.*, lipoproteins, *e.g.*, low-density lipoproteins, *e.g.*, oxidatively modified low-density lipoproteins; (3) the ability to modulate degradation, *e.g.*, proteolytic degradation, of proteins, *e.g.*, lipoproteins, *e.g.*, low-density lipoproteins, *e.g.*, oxidatively modified low-density lipoproteins; (4) the ability to modulate, *e.g.*, increase, uptake of proteins, *e.g.*, lipoproteins, *e.g.*, low-density lipoproteins, *e.g.*, oxidatively modified low-density lipoproteins, by cells, *e.g.*, macrophages and muscle cells, *e.g.*, smooth muscle cells; (5) the ability to modulate the function of a cell expressing LOX-1 or TANGO 402; (6) the ability to modulate the binding of a protein, *e.g.*, oxidized low-density lipoprotein (Ox-LDL), to a cell which expresses LOX-1 or TANGO 402; and (7) the ability to modulate the binding of a protein, *e.g.*, Ox-LDL, to LOX-1 or TANGO 402.

TANGO 402 nucleic acids, proteins and modulators thereof can be used to modulate *e.g.*, (1) the ability to modulate, *e.g.*, prevent, lipid deposition, *e.g.*, in arteries, and thus modulate, *e.g.*, prevent, intimal thickening; (2) the ability to modulate, *e.g.*, induce or prevent, changes in cells, *e.g.*, transformation of cells (*e.g.*, macrophages and smooth muscle cells) into foam cells and functional alteration of cells (*e.g.*, endothelial cells, *e.g.*, intimal neovascular endothelial cells); (3) the ability to bind and phagocytose cells, *e.g.*, aged and apoptotic cells; and (4) the ability to remove debris, *e.g.*, apoptotic cells, from blood vessel walls.

In another example, TANGO 402 nucleic acids, proteins and modulators thereof can be used to modulate *e.g.*, (1) the ability to modulate homeostasis, *e.g.*, vascular homeostasis, *e.g.*, by modulating, *e.g.*, preventing the impairment of, nitric oxide production; (2) the ability to modulate, *e.g.*, inhibit, the expression of molecules, *e.g.*, adhesion molecules (*e.g.*, leukocyte adhesion molecules) and growth factors (*e.g.*, smooth-muscle growth factors); (3) the ability to alter, *e.g.*, increase, expression in response to stimuli, *e.g.*, TNF, shear stress, and pathophysiological stimuli relevant to disorders (*e.g.*, atherosclerosis and inflammation).

In yet another example, TANGO 402 nucleic acids, proteins and modulators thereof can be used to modulate *e.g.*, (1) the ability to form, *e.g.*, stabilize, promote, facilitate, inhibit, or disrupt, cell-extracellular matrix interactions, *e.g.*, adhesion between cells and extracellular matrix; (2) the ability to form, *e.g.*, stabilize, promote, facilitate, inhibit, or disrupt, cell to cell and cell to blood product interaction, *e.g.*, between leukocytes and platelets or leukocytes and vascular endothelial cells; and (3) the ability to recognize large molecules, *e.g.*, carbohydrates.

In light of the fact that TANGO 402 is homologous to LOX-1, TANGO 402 nucleic acids, proteins and modulators thereof have biological activities that can also

include the ability to perform one or more of the functions of LOX-1 described, for example, in the following: Sawamura et al. (1997) *Nature.* 386:73-77; Kataoka et al. (1999) *Circulation.* 99:3110-3117; and Kita (1999) *Circulation Research.* 84:1113-1115, the contents of each of which is incorporated herein by reference in its entirety.

Moreover, due to TANGO 402's homology to LOX-1, as evidenced by the presence of similar domains and mapping coordinates between the two molecules, TANGO 402 nucleic acids, proteins and modulators thereof can be used to modulate or treat disorders in which LOX-1 plays a role, some of which are described in the following references: Sawamura et al. (1997) *Nature.* 386:73-77; Kataoka et al. (1999) *Circulation.* 99:3110-3117; and Kita (1999) *Circulation Research.* 84:1113-1115, the contents of each of which is incorporated herein by reference in its entirety.

Furthermore, TANGO 402 nucleic acids, proteins and modulators thereof can modulate or treat atherosclerosis, *e.g.*, by binding to oxidatively modified low density lipoprotein (Ox-LDL) and its lipid constituents, thus preventing lipid deposition and intimal thickening in the arteries, and thus preventing the induction of endothelial expression of leukocyte adhesion molecules and smooth-muscle growth factors (both which are implicated in atherogenesis).

In another example, TANGO 402 nucleic acids, proteins and modulators thereof modulate or treat immune related diseases and disorders. As LOX-1 is implicated in inflammation, and as LOX-1 has highest homology with the NKR-P1 family of proteins, which are involved in target-cell recognition and natural killer cell activation, TANGO 402 nucleic acids, proteins and modulators thereof can be used to diagnose disorders and/or modulate or treat inflammatory disorders such as bacterial infection, psoriasis, septicemia, cerebral malaria, inflammatory bowel disease, multiple sclerosis, arthritis (*e.g.*, rheumatoid arthritis, osteoarthritis), and allergic inflammatory disorders (*e.g.*, asthma, psoriasis), and processes. Further, TANGO 402 nucleic acids, proteins and modulators thereof can be used to identify, diagnose and/or modulate or treat immune disorders including, *e.g.*, autoimmune disorders (*e.g.*, arthritis, graft rejection (*e.g.*, allograft rejection), and T cell autoimmune disorders (*e.g.*, AIDS)) and inflammatory disorders.

TANGO 402 nucleic acids, proteins and modulators thereof be used to identify, diagnose and/or modulate or treat TNF-related disorders, as LOX-1 expression is induced by tumor necrosis factors. Such disorders include, *e.g.*, acute myocarditis, myocardial infarction, congestive heart failure, T cell disorders (*e.g.*, dermatitis, fibrosis)), differentiative and apoptotic disorders, and disorders related to angiogenesis (*e.g.*, tumor formation and/or metastasis, cancer). As LOX-1 expression is upregulated in hypertensive

rats, and as LOX-1 levels are downregulated in patients treated with ACE (angiotensin converting enzyme) inhibitors, TANGO 402 can also play a role in treating hypertension and congestive heart failure.

As both TANGO 402 has C-type lectin domains or C-type lectin-like domains, and is similar in that respect to the selectins, which are implicated in cell-cell recognition (including endothelial-leukocyte adhesion), TANGO 402 nucleic acids, proteins and modulators thereof can be used to identify, diagnose and/or modulate or treat cell adhesion or cell migration/motility related disorders. Such disorders include, *e.g.*, disorders associated with adhesion and migration of cells, *e.g.*, platelet aggregation disorders (*e.g.*, Glanzmann's thromboasthemia, which is a bleeding disorders characterized by failure of platelet aggregation in response to cell stimuli), inflammatory disorders (*e.g.*, leukocyte adhesion deficiency, which is a disorder associated with impaired migration of neutrophils to sites of extravascular inflammation), disorders associated with abnormal tissue migration during embryo development, and tumor metastasis.

As TANGO 402 has a C-type lectin domain or C-type lectin-like domain, TANGO 402 nucleic acids, proteins and modulators thereof can be used to diagnose C-type lectin disorders and/or modulate calcium-dependent carbohydrate recognition. TANGO 402 proteins exhibit homology to lectins. In light of this, TANGO 402 nucleic acids, proteins and modulators thereof can be utilized to modulate cell-cell, cell-extracellular matrix (ECM) interactions, cell adhesion, cell migration and cell signaling. TANGO 402 nucleic acids, proteins and modulators thereof can be utilized to treat and/or prevent disorders and diseases associated with aberrant cell-cell, cell-ECM interactions, cell migration, cell adhesion and cell-signaling, as well as treating and preventing tumor cell metastasis. TANGO 402 nucleic acids, proteins and modulators thereof can also be utilized to treat and/or prevent the migration of cancerous and precancerous cells (*e.g.*, tumor migration).

TANGO 402 nucleic acids, proteins and modulators thereof can also be used to modulate cell proliferation, *e.g.*, abnormal cell proliferation. Such modulation may, for example, be via modulation of one or more elements involved in signal transduction cascades.

TANGO 402 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, fetal tissues) and/or cells (*e.g.*, fetal cells) in which TANGO 402 is expressed. TANGO 402 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

<u>Human MANGO 346</u>

A MANGO 346 cDNA was identified from clones present in a human brain library among sequences that encode signal peptides. This analysis led to the identification of a clone, jlhbab575g04, encoding full-length human MANGO 346. The human MANGO 346 cDNA of this clone is 1196 nucleotides long (Figure 29; SEQ ID NO:123). The open reading frame of this cDNA, nucleotides 319 to 498 of SEQ ID NO:123 (SEQ ID NO:124), encodes a 60 amino acid secreted protein (Figure 18; SEQ ID NO:125).

Figure 30 depicts a hydropathy plot of human MANGO 346. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence (amino acids 1 to 19 of SEQ ID NO:125; SEQ ID NO:126) on the left from the mature protein (amino acids 20 to 60 of SEQ ID NO:125; SEQ ID NO:127) on the right.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human MANGO 346 includes a 19 amino acid signal peptide (amino acid 1 to amino acid 19 of SEQ ID NO:125; SEQ ID NO:126) preceding the mature human protein (corresponding to amino acid 20 to amino acid 60 of SEQ ID NO:125; SEQ ID NO:127). The molecular weight of protein without post-translational modifications is 7.1 kDa prior to the cleavage of the signal peptide, 5.0 kDa after cleavage of the signal peptide.

In one embodiment of a nucleotide sequence of human MANGO 346, the nucleotide at position 13 is cytosine (C)(SEQ ID NO:124). In this embodiment, the amino acid at position 5 is leucine (L)(SEQ ID NO:125). In an alternative embodiment, a species variant of human MANGO 346 has a nucleotide at position 13 which is adenine (A)(SEQ ID NO:216). In this embodiment, the amino acid at position 5 is isoleucine (I)(SEQ ID NO:217), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human MANGO 346, the nucleotide at position 59 is adenine (A)(SEQ ID NO:124). In this embodiment, the amino acid at position 20 is tyrosine (Y)(SEQ ID NO:125). In an alternative embodiment, a species variant of human MANGO 346 has a nucleotide at position 59 which is thymidine (T)(SEQ ID NO:218). In this embodiment, the amino acid at position 20 is phenylalanine (F)(SEQ ID NO:219), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human MANGO 346, the nucleotide at position 61 is thymidine (T)(SEQ ID NO:124). In this embodiment, the amino acid at position 21 is serine (S)(SEQ ID NO:125). In an alternative embodiment, a species variant of human MANGO 346 has a nucleotide at position 61 which is adenine

(A)(SEQ ID NO:220). In this embodiment, the amino acid at position 21 is threonine (T)(SEQ ID NO:221), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human MANGO 346, the nucleotide at position 80 is guanine (G)(SEQ ID NO:124). In this embodiment, the amino acid at position 27 is arginine (R)(SEQ ID NO:125). In an alternative embodiment, a species variant of human MANGO 346 has a nucleotide at position 80 which is adenine (A)(SEQ ID NO:222). In this embodiment, the amino acid at position 27 is lysine (K)(SEQ ID NO:223), *i.e.*, a conservative substitution.

One protein kinase C phosphorylation site is present in human MANGO 346 which has the sequence, TIK (at amino acids 44 to 46 of SEQ ID NO:125). Human MANGO 346 has three Casein Kinase II phosphorylation sites. The first has the sequence SFLE (at amino acids 21 to 24 of SEQ ID NO:125), the second has the sequence TIKE (at amino acids 44 to 47 of SEQ ID NO:125) and the third has the sequence TYYD (at amino acids 51 to 54 of SEQ ID NO:125). Human MANGO 346 has one prokaryotic membrane lipoprotein lipid attachment site. The sequence is CILPLLLLASC (at amino acids 6 to 16 of SEQ ID NO:125).

Clone EpM346, which encodes human MANGO 346, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-291. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

<u>Uses of MANGO 346 Nucleic acids, Polypeptides, and Modulators Thereof</u>

As MANGO 346 was originally found in a human brain library, nucleic acids, proteins, and modulators thereof can be used to diagnose or identify disorders and/or modulate the proliferation, development, differentiation, and/or function of neural organs, *e.g.*, neural tissues and cells, *e.g.*, cells of the central nervous system, *e.g.*, cells of the peripheral nervous system. MANGO 346 nucleic acids, proteins, and modulators thereof can also be used to diagnose or identify disorders and/or modulate symptoms associated with abnormal neural signaling and function, *e.g.*, epilepsy, spinal cord injuries, infarction, infection, malignancy, exposure to toxic agents, nutritional deficiency, paraneoplastic syndromes, and degenerative nerve diseases including but not limited to Alzheimer's disease, Parkinson's disease, Huntington's Chorea, amyotrophic lateral sclerosis, progressive supra-nuclear palsy, and other dementias.

MANGO 346 nucleic acids, proteins and modulators thereof can, in addition to the above, be utilized to diagnose disorders, regulate or modulate development and/or differentiation of processes involved in central or peripheral nervous system formation and activity, as well as in ameliorating any symptom associated with a disorder of such cell types, tissues and organs.

MANGO 346 nucleic acids, proteins and modulators thereof can, in addition to the above, be utilized to regulate or diagnose disorders, modulate development and/or differentiation of processes involved in central or peripheral nervous system formation and activity, as well as in ameliorating any symptom associated with a disorder of such cell types, tissues and organs. Furthermore, the TANGO 346 proteins can be used to disrupt protein interaction or cellular signaling in brain tissues or cells. In particular, TANGO 346 proteins are useful to treat neural related disorders or neural damage, such as for regenerative neural repair after damage by trauma, degeneration, or inflammation *e.g.*, spinal cord injuries, infarction, infection, malignancy, exposure to toxic agents, nutritional deficiency, paraneoplastic syndromes, and degenerative nerve diseases including but not limited to Alzheimer's disease, Parkinson's disease, Huntington's Chorea, amyotrophic lateral sclerosis, progressive supra-nuclear palsy, and other dementias.

As MANGO 346 is a secreted protein and thus likely a signaling molecule, MANGO 346 nucleic acids, proteins or modulators thereof, can be used to modulate MANGO 346 biological activities, which include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.*, in receptor-ligand recognition; (2) ability to modulate cell-cell interactions; (3) the ability to modulate the proliferation, differentiation and/or activity of neural cells; and (4) the ability to modulate intracellular signaling cascades (*e.g.*, signal transduction cascades).

MANGO 346 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the brain) and/or cells (*e.g.*, neurons) in which MANGO 346 is expressed. MANGO 346 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

Human MANGO 349

A cDNA encoding human MANGO 349 was identified by analyzing the sequences of clones present in a human brain library for sequences that encode wholly secreted or transmembrane proteins. This analysis led to the identification of a clone, jlhbae318gd08, encoding full-length human MANGO 349. The human cDNA of this clone is 3649 nucleotides long (Figure 31; SEQ ID NO:128). The open reading frame of this cDNA,

nucleotides 221 to 721 of SEQ ID NO:128 (SEQ ID NO:129), encodes a 167 amino acid secreted protein (Figure 31; SEQ ID NO:130).

Figure 32 depicts a hydropathy plot of human MANGO 349. Relatively hydrophobic regions of the protein are above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The dashed vertical line separates the signal sequence from the mature protein described below.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human MANGO 349 includes a 26 amino acid signal peptide (amino acid 1 to amino acid 26 of SEQ ID NO:130; SEQ ID NO:131) preceding the mature human protein (corresponding to amino acid 27 to amino acid 167 of SEQ ID NO:130; SEQ ID NO:132). The molecular weight of human protein without post-translational modifications is 17.6 kDa prior to the cleavage of the signal peptide, 15.1 kDa after cleavage of the signal peptide.

In one embodiment of a nucleotide sequence of human MANGO 349, the nucleotide at position 4 is adenine (A)(SEQ ID NO:129). In this embodiment, the amino acid at position 2 is threonine (T)(SEQ ID NO:130). In an alternative embodiment, a species variant of human MANGO 349 has a nucleotide at position 4 which is thymine (T)(SEQ ID NO:224). In this embodiment, the amino acid at position 2 is serine (S)(SEQ ID NO:225), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human MANGO 349, the nucleotide at position 61 is adenine (A)(SEQ ID NO:129). In this embodiment, the amino acid at position 21 is isoleucine (I)(SEQ ID NO:130). In an alternative embodiment, a species variant of human MANGO 349 has a nucleotide at position 61 which is cytosine (C)(SEQ ID NO:226). In this embodiment, the amino acid at position 21 is leucine (L)(SEQ ID NO:227), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human MANGO 349, the nucleotide at position 86 is guanine (G)(SEQ ID NO:129). In this embodiment, the amino acid at position 29 is arginine (R)(SEQ ID NO:130). In an alternative embodiment, a species variant of human MANGO 349 has a nucleotide at position 86 which is adenine (A)(SEQ ID NO:228). In this embodiment, the amino acid at position 29 is lysine (K)(SEQ ID NO:229), *i.e.*, a conservative substitution.

In one embodiment of a nucleotide sequence of human MANGO 349, the nucleotide at position 123 is guanine (G)(SEQ ID NO:129). In this embodiment, the amino acid at position 41 is glutamate (E)(SEQ ID NO:130). In an alternative embodiment, a species variant of human MANGO 349 has a nucleotide at position 123

which is cytosine (C)(SEQ ID NO:230). In this embodiment, the amino acid at position 41 is aspartate (D)(SEQ ID NO:231), *i.e.*, a conservative substitution.

Two Protein C Kinase phosphorylation sites are present in human MANGO 349. The first has the sequence SLK (at amino acids 136 to 139 of SEQ ID NO:130) and the second has the sequence SGR (at amino acids 152 to 154 of SEQ ID NO:130). Two casein kinase II phosphorylation sites are present in human MANGO 349. The first has the sequence SGTE (at amino acids 38 to 41 of SEQ ID NO:130), and the second has the sequence SGRE (at amino acids 152 to 155 of SEQ ID NO:130). Human MANGO 349 has four N-myristylation sites. The first has the sequence GGILAT (at amino acids 10 to 15 of SEQ ID NO:130), the second has the sequence GTEVAD (at amino acids 39 to 44 of SEQ ID NO:130), the third has the sequence GVAASH (at amino acids 89 to 94 of SEQ ID NO:17), and the fourth has the sequence GGPPSL (at amino acids 132 to 137 of SEQ ID NO:130).

Clone EpM349, which encodes human MANGO 349, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on June 29, 1999 and assigned Accession Number PTA-295. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

<u>Uses of MANGO 349 Nucleic acids, Polypeptides, and Modulators Thereof</u>

As MANGO 349 was originally found in a human brain library, nucleic acids, proteins, and modulators thereof can be used to diagnose or identify disorders and/or modulate the proliferation, development, differentiation, and/or function of neural organs, *e.g.*, neural tissues and cells, *e.g.*, cells of the central nervous system, *e.g.*, cells of the peripheral nervous system. MANGO 349 nucleic acids, proteins, and modulators thereof can also be used to diagnose or identify disorders and/or modulate symptoms associated with abnormal neural signaling and function, *e.g.*, epilepsy, stroke, traumatic injury, etc.

MANGO 349 nucleic acids, proteins and modulators thereof can, in addition to the above, be utilized to diagnose disorders, regulate or modulate development and/or differentiation of processes involved in central or peripheral nervous system formation and activity, as well as in ameliorating any symptom associated with a disorder of such cell types, tissues and organs. Furthermore, the TANGO 349 proteins can be used to disrupt protein interaction or cellular signaling in brain tissues or cells. In particular, TANGO 349 proteins could be useful to treat neural related disorders or neural damage, such as for

regenerative neural repair after damage by trauma, degeneration, or inflammation *e.g.*, spinal cord injuries, infarction, infection, malignancy, exposure to toxic agents, nutritional deficiency, paraneoplastic syndromes, and degenerative nerve diseases including but not limited to Alzheimer's disease, Parkinson's disease, Huntington's Chorea, amyotrophic lateral sclerosis, progressive supra-nuclear palsy, and other dementias.

As MANGO 349 is a secreted protein and thus likely a signaling molecular, MANGO 349 nucleic acids, proteins and modulators thereof can be used to diagnose disorders and/or modulate MANGO 349 biological activities, which include, *e.g.*, (1) the ability to modulate, *e.g.*, stabilize, promote, inhibit or disrupt, protein-protein interactions (*e.g.*, homophilic and/or heterophilic), and protein-ligand interactions, *e.g.* in receptor-ligand recognition; (2) ability to modulate cell-cell interactions; (3) the ability to modulate proliferation, differentiation and/or activity of neural cells; and (4) the ability to modulate intracellular signaling cascades (*e.g.* signal transduction cascades).

MANGO 349 expression can be utilized as a marker (*e.g.*, an *in situ* marker) for specific tissues (*e.g.*, the brain) and/or cells (*e.g.*, neurons) in which MANGO 349 is expressed. MANGO 349 nucleic acids can also be utilized for chromosomal mapping, or as chromosomal markers, *e.g.*, in radiation hybrid mapping.

Tables 1 provides a summary of the sequence information for TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, human TANGO 393, mouse TANGO 393, TANGO 402, MANGO 346, and MANGO 349.

Table 2 provides a summary of the domains of TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, human TANGO 393, mouse TANGO 393, TANGO 402, MANGO 346, and MANGO 349. It is noted that human and mouse TANGO 393 leucine-rich repeats are not included in Table 2, but are described *supra*.

TABLE 1: Summary of Human TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, MANGO 349 and Mouse TANGO 393 Sequence Information

Gene	cDNA	ORF	Figure	Accession Number
Human TANGO 339	SEQ ID NO:1	SEQ ID NO:2	Figure 1	PTA-292
Human TANGO 353	SEQ ID NO:27	SEQ ID NO:28	Figure 5	PTA-292
Human TANGO 358	SEQ ID NO:36	SEQ ID NO:37	Figure 7	PTA-292
Human TANGO 365	SEQ ID NO:44	SEQ ID NO:45	Figure 9	PTA-291
Human TANGO 368	SEQ ID NO:52	SEQ ID NO:53	Figure 11	PTA-291
Human TANGO 369	SEQ ID NO:58	SEQ ID NO:59	Figure 14	PTA-295
Human TANGO 383	SEQ ID NO:63	SEQ ID NO:64	Figure 16	PTA-295
Human TANGO 393	SEQ ID NO:73	SEQ ID NO:74	Figure 19	PTA-295
Mouse TANGO 393	SEQ ID NO:93	SEQ ID NO:94	Figure 21	
Human TANGO 402	SEQ ID NO:110	SEQ ID NO:111	Figure 25	PTA-294
Human MANGO 346	SEQ ID NO:123	SEQ ID NO:124	Figure 29	PTA-291
Human MANGO 349	SEQ ID NO:128	SEQ ID NO:129	Figure 31	PTA-295

TABLE 2: Summary of Domains of TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 Proteins

Protein	Signal Sequence	Mature Protein	Extracellular	Transmembrane 4-like	Peripherin/ rom-1-like	C-type Lectin-like	Transmembrane	Cytoplasmic
HUMAN TANGO 339	aa 1-42 of SEQ ID NO:3 (SEQ ID NO:5)	aa 43-270 of SEQ ID NO:3 (SEQ ID NO:4)	aa 43-61 and 116-232 of SEQ ID NO:3 (SEQ ID NOs:20 and 21)	aa 68-260 of SEQ ID NO:3 (SEQ ID NO:6)	aa 18-270 of SEQ ID NO:3 (SEQ ID NO:8)		aa 62-84; 93-115 and 233-254 of SEQ ID NO:3 (SEQ ID NO:15, 16, 17)	aa 85-92 and 255-270 of SEQ ID NO:3 (SEQ ID NO:22 and 23)
HUMAN TANGO 353	aa 1-14 of SEQ ID NO:29 (SEQ ID NO:31)	aa 15-230 of SEQ ID NO:29 (SEQ ID NO:30)	aa 15-116 of SEQ ID NO:29 (SEQ ID NO:32)				aa 117-141 of SEQ ID NO:29 (SEQ ID NO:33)	aa 142-230 of SEQ ID NO:29 (SEQ ID NO:34)
HUMAN TANGO 358	aa 1-42 of SEQ ID NO:38 (SEQ ID NO:40)	aa 43-82 of SEQ ID NO:38 (SEQ ID NO:39)	aa 43-49 of SEQ ID NO:38 (SEQ ID NO:41)				aa 50-66 of SEQ ID NO:38 (SEQ ID NO:42)	aa 67-82 of SEQ ID NO:38 (SEQ ID NO:43)
HUMAN TANGO 365	aa 1-36 of SEQ ID NO:46; (SEQ ID NO:47)	aa 37-165 of SEQ ID NO:46; (SEQ ID NO:48)	aa 95-165 of SEQ ID NO:46 (SEQ ID NO:51)				aa 52-70 and aa 78-94 of SEQ ID NO:46; (SEQ ID NO:49 and SEQ ID NO:50)	aa 71-77 of SEQ ID NO:46; (SEQ ID NO:232)
HUMAN TANGO 368	aa 1-26 of SEQ ID NO:54 (SEQ ID NO:56)	aa 27-59 of SEQ ID NO:54 (SEQ ID NO:55)						

Protein	Signal Sequence	Mature Protein	Extracellular	Transmembrane 4-like	Peripherin/ rom-1-like	C-type Lectin-like	Transmembrane	Cytoplasmic
HUMAN TANGO 369	aa 1-26 of SEQ ID NO:60 (SEQ ID NO:62)	aa 27-58 of SEQ ID NO:60 (SEQ ID NO:61)						
HUMAN TANGO 383	aa 1-20 of SEQ ID NO:65; (SEQ ID NO:66)	aa 21-140 of SEQ ID NO:65; (SEQ ID NO:67)	aa 21-49 and aa 134-140 of SEQ ID NO:65 (SEQ ID NOs: 233 and 136)				aa 50-70 and aa 116-133 of SEQ ID NO:65; (SEQ ID NO:68 and SEQ ID NO:69)	aa 71-115 of SEQ ID NO:65; (SEQ ID NO:70)
HUMAN TANGO 393	aa 1-26 of SEQ ID NO:75; (SEQ ID NO:76)	aa 27-473 of SEQ ID NO:75; (SEQ ID NO:77)	aa 27-447 of SEQ ID NO:75 (SEQ ID NO:89)				aa 448-467 of SEQ ID NO:75; (SEQ ID NO:78)	aa 468-473 of SEQ ID NO:75; (SEQ ID NO:134)
MOUSE TANGO 393	aa 1-26 of SEQ ID NO:95; (SEQ ID NO:96)	aa 27-473 of SEQ ID NO:95; (SEQ ID NO:97)	aa 27-449 of SEQ ID NO:95 (SEQ ID NO:109)				aa 450-467 and aa 448-467 of SEQ ID NO:95; (SEQ ID NO:98 and SEQ ID NO:99)	aa 468-473 of SEQ ID NO:95; (SEQ ID NO:135)
HUMAN TANGO 402	aa 1-50 of SEQ ID NO:112 (SEQ ID NO:114)	aa 51-207 of SEQ ID NO:112 (SEQ ID NO:113)	aa 51-133 of SEQ ID NO:112 (SEQ ID NO:115)			aa 104-193 of SEQ ID NO:112 (SEQ ID NO:118)	aa 134-151 of SEQ ID NO:112 (SEQ ID NO:116)	aa 152-207 of SEQ ID NO:112 (SEQ ID NO:117)

Protein	Signal Sequence	Mature Protein	Extracellular	Transmembrane 4-like	Peripherin/ rom-1-like	C-type Lectin-like	Transmembrane	Cytoplasmic
HUMAN MANGO 346	aa 1-19 of SEQ ID NO:125; (SEQ ID NO:126)	aa 20-60 of SEQ ID NO:125; (SEQ ID NO:127)						
HUMAN MANGO 349	aa 1-26 of SEQ ID NO:130; (SEQ ID NO:131)	aa 27-167 of SEQ ID NO:130; (SEQ ID NO:132)						

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a biologically active portion thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. As used herein, the term "isolated"when referring to a nucleic acid molecule does not include an isolated chromosome.

In instances wherein the nucleic acid molecule is a cDNA or RNA, *e.g.*, mRNA, molecule, such molecules can include a poly A "tail", or, alternatively, can lack such a 3' tail. Although cDNA or RNA nucleotide sequences may be depicted herein with such tail sequences, it is to be understood that cDNA nucleic acid molecules of the invention are also intended to include such sequences lacking the depicted poly A tails.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220,

222, 224, 226, 228 or 230 or a complement thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230 as a hybridization probe, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or the nucleotide sequence of the cDNA of a clone deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence under the conditions set forth herein, thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding a full length polypeptide of the invention for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a polypeptide of the invention. The nucleotide sequence determined from the cloning one gene allows for the generation of probes and primers designed for use in identifying and/or cloning homologs in other cell types, *e.g.*, from other tissues, as well as

homologs from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or the nucleotide sequence of the cDNA of a clone deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, or of a naturally occurring mutant of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, or 129. In another embodiment, the oligonucleotide comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least 400, preferably 450, 500, 530, 550, 600, 700,800, 900, 1000 or 1150 consecutive oligonucleotides of the sense or antisense sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or the nucleotide sequence of the cDNA of a clone deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, or of a naturally occurring mutant of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, or 129.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences encoding the same protein molccule encoded by a selected nucleic acid molecule. The probe comprises a label group attached thereto, *e.g.*, a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, *e.g.*, detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

A nucleic acid fragment encoding a biologically active portion of a polypeptide of the invention can be prepared by isolating a portion of any of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229

or 231, or the nucleotide sequence of the cDNA of a clone deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, expressing the encoded portion of the polypeptide protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the polypeptide.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or the nucleotide sequence of the cDNA of a clone deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231.

In addition to the nucleotide sequences of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or the nucleotide sequence of the cDNA of a clone deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence may exist within a population (*e.g.*, the human population). Such genetic polymorphisms may exist among individuals within a population due to natural allelic variation.

An allele is one of a group of genes which occur alternatively at a given genetic locus. For example, TANGO 393 has been mapped to chromosome 22, and therefore TANGO 393 family members can include nucleotide sequence polymorphisms (*e.g.*, nucleotide sequences that vary from SEQ ID NO:73 and SEQ ID NO:74) that map to this chromosome 22 region, and such sequences represent TANGO 393 allelic variants. In another example, TANGO 339 has been mapped to chromosome 10, and therefore TANGO 339, family members can include nucleotide sequence polymorphisms (*e.g.*, nucleotide sequences that vary from SEQ ID NO:1 and SEQ ID NO:2) that map to this chromosome 10 region, and such sequences represent allelic variants.

As used herein, the phrase "allelic variant" refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention. In one embodiment, polymorphisms that are associated with a particular disease and/or disorder are used as markers to diagnose said disease or disorder. In a preferred embodiment, polymorphisms are used as a marker to diagnose abnormal coronary function such as atherosclerosis.

Moreover, nucleic acid molecules encoding proteins of the invention from other species (homologs), which have a nucleotide sequence which differs from that of the human or mouse protein described herein are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologs of a cDNA of the invention can be isolated based on their identity to the human nucleic acid molecule disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble form of a membrane-bound protein of the invention isolated based on its hybridization to a nucleic acid molecule encoding all or part of the membrane-bound form. Likewise, a cDNA encoding a membrane-bound form can be isolated based on its hybridization to a nucleic acid molecule encoding all or part of the soluble form.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400. 1500, 1600, 1700, 1800, 1900 or 2000 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:1, 27, 36, 44, 52, 58, 63, 73, 93, 110, 123, or 128 or a complement thereof.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 20, 50, 100, 200, 300, 400, 500, 600, 700, 800 or 900 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising

the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:2, 28, 37, 45, 53, 59, 64, 74, 94, 111, 124, or, 129, or a complement thereof.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45° C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65° C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or a complement thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologs of various species may be non-essential for activity and thus would be likely targets for alteration. Specific examples of conservative amino acid alterations from the original sequence are shown in SEQ ID NOs:137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231. Alternatively, amino acid residues that are conserved among the orthologs of various species (*e.g.*, murine and human) may be essential for activity and thus would not be likely targets for alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or complement thereof, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid, asparagine, glutamine), uncharged polar side chains (*e.g.*, glycine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify

mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant polypeptide that is a variant of a polypeptide of the invention can be assayed for: (1) the ability to form protein-protein interactions with proteins in a signaling pathway of the polypeptide of the invention such as in central nervous system cells, lymphoid cells, hypothalamus cells, or prostate cells with the proteins encoded by the genes of the present invention (*e.g.*, leucine-rich repeat interactions and transmembrane 4); (2) the ability to bind a ligand of the polypeptide of the invention (*i.e.*, in transmembrane proteins of the invention or alternatively, secreted proteins which are the ligand for a cellular receptor); or (3) the ability to bind to an intracellular target protein of the polypeptide of the invention. In yet another preferred embodiment, the mutant polypeptide can be assayed for the ability to modulate cellular proliferation, cellular migration or chemotaxis, or cellular differentiation.

The present invention encompasses antisense nucleic acid molecules, *i.e.*, molecules which are complementary to a sense nucleic acid encoding a polypeptide of the invention, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, *e.g.*, all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides or more in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil,

dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et

al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of the invention can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (*e.g.*, the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al. (1996) *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup (1996), *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996), *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra*, and Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al. (1989) *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W0 88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.*, Krol et al. (1988) *Bio/Techniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide

may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide of the invention. In one embodiment, the native polypeptide can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, polypeptides of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide of the invention can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, *i.e.*, it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein (*e.g.*, the amino acid sequence shown in any of SEQ ID NO:6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 32, 33, 34, 41, 42, 43, 49, 50, 51, 68, 69, 70, 71, 72, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 115, 116, 117, 118, 119, 133, 134, 135, or 136, which include fewer amino acids than the full length protein, and exhibit at least one activity of

the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Preferred polypeptides have the amino acid sequence of SEQ ID NO:6, 7, 8, 9, 10, 19, 20, 21, 22 or 23. Other useful proteins are substantially identical (*e.g.*, at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99%) to any of SEQ ID NO:6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 32, 33, 34, 41, 42, 43, 49, 50, 51, 68, 69, 70, 71, 72, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 115, 116, 117, 118, 119, 133, 134, 135, or 136, and retain the functional activity of the protein of the corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions (*e.g.*, overlapping positions) x 100). In one embodiment, the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences

homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov.

Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti (1994) *Comput. Appl. Biosci., 10*:3-5; and FASTA described in Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. 85*:2444-8. Within FASTA, ktup is a control option that sets the sensitivity and speed of the search. If ktup=2, similar regions in the two sequences being compared are found by looking at pairs of aligned residues; if ktup=1, single aligned amino acids are examined. ktup can be set to 2 or 1 for protein sequences, or from 1 to 6 for DNA sequences. The default if ktup is not specified is 2 for proteins and 6 for DNA. For a further description of FASTA parameters, see http://bioweb.pasteur.fr/docs/man/man/fasta.1.html#sect2, the contents of which are incorporated herein by reference.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably biologically active) of a polypeptide of the invention operably linked to a heterologous polypeptide (*i.e.*, a polypeptide other than the same polypeptide of the invention). Within the fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the N-terminus or C-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused to the C-terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus. For example, the native signal sequence of a polypeptide of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction may be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel et al., *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

A signal sequence of a polypeptide of the invention (SEQ ID NO:5, 31, 40, 47, 56, 62, 66, 76, 96, 114, 126 or 131) can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal sequence, as well as to the signal sequence itself and to the polypeptide in the absence of the signal sequence (*i.e.*, the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal sequence of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

In another embodiment, the signal sequences of the present invention can be used to identify regulatory sequences, *e.g.*, promoters, enhancers, repressors. Since signal sequences are the most amino-terminal sequences of a peptide, it is expected that the nucleic acids which flank the signal sequence on its amino-terminal side will be regulatory sequences which affect transcription. Thus, a nucleotide sequence which encodes all or a portion of a signal sequence can be used as a probe to identify and isolate signal sequences and their flanking regions, and these flanking regions can be studied to identify regulatory elements therein.

The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring

form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the protein of the invention for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (*see, e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries,

can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA 89*:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

The polypeptides of the invention can exhibit post-translational modifications, including, but not limited to glycosylations, (*e.g.*, N-linked or O-linked glycosylations), myristylations, palmitylations, acetylations and phosphorylations (*e.g.*, serine/threonine or tyrosine). In one embodiment, the TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 polypeptides of the invention exhibit reduced levels of O-linked glycosylation and/or N-linked glycosylation relative to endogenously expressed TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, MANGO 349 polypeptides. In another embodiment, the TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 or MANGO 349 polypeptides of the invention do not exhibit O-linked glycosylation or N-linked glycosylation.

The polypeptides of the invention can, for example, include modifications that can increase such attributes as stability, half-life, ability to enter cells and aid in administration, *e.g.*, in vivo administration of the polypeptides of the invention. For example, polypeptides of the invention can comprise a protein transduction domain of the HIV TAT protein as described in Schwarze, et al. (1999 *Science* 285:1569-1572), thereby facilitating delivery of polypeptides of the invention into cells.

An isolated polypeptide of the invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, *e.g.*, hydrophilic regions. Figures 2, 6, 8, 10, 12, 15, 17, 20, 22, 26, 30, and 32 are hydropathy plots of the proteins of the invention. These

plots or similar analyses can be used to identify hydrophilic regions. In certain embodiments, the nucleic acid molecules of the invention are present as part of nucleic acid molecules comprising nucleic acid sequences that contain or encode heterologous (*e.g.*, vector, expression vector, or fusion protein) sequences. These nucleotides can then be used to express proteins which can be used as immunogens to generate an immune response, or more particularly, to generate polyclonal or monoclonal antibodies specific to the expressed protein.

An immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal). An appropriate immunogenic preparation can contain, for example, recombinantly expressed or chemically synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention, *e.g.*, an epitope of a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. Preferred polyclonal antibody compositions are ones that have been selected for antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred polyclonal antibody preparations are ones that contain only antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen compositions made using a non-human host cell for recombinant expression of a polypeptide of the invention. In such a manner, the only human epitope or epitopes

recognized by the resulting antibody compositions raised against this immunogen will be present as part of a polypeptide or polypeptides of the invention.

The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be isolated from the mammal (*e.g.*, from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. Alternatively, antibodies specific for a protein or polypeptide of the invention can be selected for (*e.g.*, partially purified) or purified by, *e.g.*, affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention is produced as described herein, and covalently or non-covalently coupled to a solid support such as, for example, a chromatography column. The column can then be used to affinity purify antibodies specific for the proteins of the invention from a sample containing antibodies directed against a large number of different epitopes, thereby generating a substantially purified antibody composition, *i.e.*, one that is substantially free of contaminating antibodies. By a substantially purified antibody composition is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than those on the desired protein or polypeptide of the invention, and preferably at most 20%, yet more preferably at most 10%, and most preferably at most 5% (by dry weight) of the sample is contaminating antibodies. A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein or polypeptide of the invention.

At an appropriate time after immunization, *e.g.*, when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (*see generally Current Protocols in Immunology* (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage

display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J.* 12:725-734.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, *e.g.*, Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816,397, which are incorporated herein by reference in their entirety.) Humanized antibodies are antibody molecules from non-human species having one or more complementarily determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. (See, *e.g.*, Queen, U.S. Patent No. 5,585,089, which is incorporated herein by reference in its entirety.) Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison (1985) *Science* 229:1202-1207; Oi et al. (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, *see, e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. (Jespers et al. (1994) *Bio/technology* 12:899-903).

An antibody directed against a polypeptide of the invention (*e.g.*, monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include

umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Further, an antibody (or fragment thereof) can be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α-interferon, β-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator, a thrombotic agent or an anti-angiogenic agent, *e.g.*, angiostatin or endostatin; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), interleukin-10 ("IL-10"), interleukin-12 ("IL-12"), interferon-γ ("IFN-γ"), interferon-α ("IFN-α"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other immune or growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld et al. (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For Drug Delivery", in

Controlled Drug Delivery (2nd Ed.), Robinson et al. (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera et al. (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin et al. (eds.), pp. 303-16 (Academic Press 1985), and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982).

Alternatively, an antibody of the invention can be conjugated to a second antibody to form an "antibody heteroconjugate" as described by Segal in U.S. Patent No. 4,676,980 or alternatively, the antibodies can be conjugated to form an "antibody heteropolymer" as described in Taylor *et al.*, in U.S. Patent Nos. 5,470,570 and 5,487,890.

An antibody with or without a therapeutic moiety conjugated to it can be used as a therapeutic that is administered alone or in combination with cytotoxic factor(s) and/or cytokine(s).

In yet a further aspect, the invention provides substantially purified antibodies or fragments thereof, including human or non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide of the invention comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295; a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295; an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession

Number PTA 291, PTA 292, PTA 294 or PTA 295, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or to the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or fragments thereof, can be human, non-human, chimeric and/or humanized antibodies.

In another aspect, the invention provides human or non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295; a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295; an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4;

and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230 or to the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide of the invention comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295; a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295; an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which

hybridizes to the nucleic acid molecule consisting of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230 or the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

The substantially purified antibodies or fragments thereof specifically bind to a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a cytoplasmic domain cytoplasmic membrane of a polypeptide of the invention. In a particularly preferred embodiment, the substantially purified antibodies or fragments thereof, the non-human antibodies or fragments thereof, and/or the monoclonal antibodies or fragments thereof, of the invention specifically bind to a secreted sequence, or alternatively, to an extracellular domain of the amino acid sequence of the invention. Examples of extracellular domains of the invention are shown in SEQ ID NOs:20, 21, 32, 41, 51, 89, 109, 112, 115, 136 or 233.

Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Still another aspect of the invention is a method of making an antibody that specifically recognizes TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349, the method comprising immunizing a mammal with a polypeptide. The polypeptide used as an immunogen comprises an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159,

161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or an amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession Number PTA 291, PTA 292, PTA 294 or PTA 295; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NOs:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, or the cDNA of a clone deposited as ATCC® Accession Number PTA 291, PTA 292, PTA 294 or PTA 295, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. After immunization, a sample is collected from the mammal that contains an antibody that specifically recognizes the immunogen. Preferably, the polypeptide is recombinantly produced using a non-human host cell. Optionally, the antibodies can be further purified from the sample using techniques well known to those of skill in the art. The method can further comprise producing a monoclonal antibody- producing cell from the cells of the mammal. Optionally, antibodies are collected from the antibody-producing cell.

III. <u>Recombinant Expression Vectors and Host Cells</u>

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide of the invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which

refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (*e.g.*, *E. coli*) or eukaryotic cells (*e.g.*, insect cells (using baculovirus expression vectors), yeast cells or mammalian

cells). Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al., *supra*.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which

allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (*Reviews - Trends in Genetics*, Vol. 1(1) 1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (*e.g.*, *E. coli*) or eukaryotic cell (*e.g.*, insect cells, yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid

can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

In another embodiment, the expression characteristics of an endogenous (*e.g.*, TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346 and MANGO 349) nucleic acid within a cell, cell line or microorganism may be modified by inserting a DNA regulatory element heterologous to the endogenous gene of interest into the genome of a cell, stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous gene (*e.g.*, TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 genes) and controls, modulates or activates the endogenous gene. For example, endogenous TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 genes which are normally "transcriptionally silent", *i.e.*, a TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 genes which are normally not expressed, or are expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, transcriptionally silent, endogenous TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 genes may be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with and activates expression of endogenous TANGO 339, TANGO 353, TANGO 358, TANGO, 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 349 genes, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide of the invention using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another

embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which sequences encoding a polypeptide of the invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a polypeptide of the invention have been introduced into their genome or homologous recombinant animals in which endogenous encoding a polypeptide of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide and for identifying and/or evaluating modulators of polypeptide activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing nucleic acid encoding a polypeptide of the invention (or a homologue thereof) into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986) and Wakayama *et al.*, (1999), *Proc. Natl. Acad. Sci. USA*, 96:14984-14989. Similar methods are used for production of other

transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide of the invention into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (*see, e.g.*, Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (*see, e.g.*, Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (*see, e.g.*, Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One

example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, *see, e.g.,* Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) *Nature* 385:810-813 and PCT Publication NOS. WO 97/07668 and WO 97/07669.

IV. Pharmaceutical Compositions

The nucleic acid molecules, polypeptides, and antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention and one or more additional active compounds.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound

into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable,

biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank et al. ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

Antibodies or antibodies conjugated to therapeutic moieties can be administered to an individual alone or in combination with cytotoxic factor(s), chemotherapeutic drug(s), and/or cytokine(s). If the latter, preferably, the antibodies are administered first and the cytotoxic factor(s), chemotherapeutic drug(s) and/or cytokine(s) are administered thereafter within 24 hours. The antibodies and cytotoxic factor(s), chemotherapeutic drug(s) and/or cytokine(s) can be administered by multiple cycles depending upon the clinical response of the patient. Further, the antibodies and cytotoxic factor(s), chemotherapeutic drug(s) and/or cytokine(s) can be administered by the same or separate routes, for example, by intravenous, intranasal or intramuscular administration. Cytotoxic factors include, but are not limited to, TNF-α, TNF-β, IL-1, IFN-γ and IL-2.

Chemotherapeutic drugs include, but are not limited to, 5-fluorouracil (5FU), vinblastine, actinomycin D, etoposide, cisplatin, methotrexate and doxorubicin. Cytokines include, but are not limited to, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10 and IL-12.

As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e*,. including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect

which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (*see, e.g.,* Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. <u>Uses and Methods of the Invention</u>

The nucleic acid molecules, proteins, protein homologs, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (*e.g.*, chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (*e.g.*, therapeutic and prophylactic).

For example, polypeptides of the invention can to used to (i) modulate cellular proliferation; (ii) modulate cellular differentiation; and/or (iii) modulate cellular adhesion. The isolated nucleic acid molecules of the invention can be used to express proteins (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (*e.g.*, in a biological sample) or a genetic lesion, and to modulate activity of a polypeptide of the invention. In addition, the polypeptides of the invention can be used to screen drugs or compounds which modulate activity or expression of a polypeptide of the invention as well as to treat disorders characterized by insufficient or excessive production of a protein of the invention or production of a form of a protein of the invention which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies of the invention can be used to detect and isolate a protein of the and modulate activity of a protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to polypeptide of the invention or have a stimulatory or inhibitory effect on, for example, expression or activity of a polypeptide of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a polypeptide of the invention or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed.*

Engl. 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Bio/Techniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent NOS. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382; and Felici (1991) *J. Mol. Biol.* 222:301-310).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to

modulate the activity of the polypeptide or a biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide protein to bind to or interact with a target molecule.

Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described above for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide (*e.g.*, a polypeptide of the invention) binds or interacts with in nature, for example, a molecule on the surface of a cell which expresses the selected protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A target molecule can be a polypeptide of the invention or some other polypeptide or protein. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*e.g.*, intracellular Ca^{2+}, diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (*e.g.*, a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises

determining the ability of the test compound to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished, for example, by determining the ability of the polypeptide to bind to a target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to further modulate the target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting a polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the polypeptide to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the polypeptide of the invention or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test

compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or A polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptide of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target molecule.

In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a candidate compound and the expression of the selected mRNA or protein (*i.e.*, the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is determined. The level of expression of the selected mRNA or protein in the presence of the candidate compound is

compared to the level of expression of the selected mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression of the polypeptide of the invention based on this comparison. For example, when expression of the selected mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of the selected mRNA or protein expression. Alternatively, when expression of the selected mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods described herein.

In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Bio/Techniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with the polypeptide of the invention and modulate activity of the polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. <u>Detection Assays</u>

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

<u>1. Chromosome Mapping</u>

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly,

nucleic acid molecules described herein or fragments thereof, can be used to map the location of the corresponding genes on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the sequence of a gene of the invention. Computer analysis of the sequence of a gene of the invention can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will yield an amplified fragment. For a review of this technique, see D'Eustachio et al. ((1983) *Science* 220:919-924).

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic acid sequences of the invention to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a gene to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes (CITE), and pre-selection by hybridization to chromosome specific cDNA libraries. Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. For a review of this technique, see Verma et al., (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988)).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can

then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland et al. (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a gene of the invention can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Furthermore, the nucleic acid sequences disclosed herein can be used to perform searches against "mapping databases", *e.g.*, BLAST-type search, such that the chromosome position of the gene is identified by sequence homology or identity with known sequence fragments which have been mapped to chromosomes.

A polypeptide and fragments and sequences thereof and antibodies specific thereto can be used to map the location of the gene encoding the polypeptide on a chromosome. This mapping can be carried out by specifically detecting the presence of the polypeptide in members of a panel of somatic cell hybrids between cells of a first species of animal from which the protein originates and cells from a second species of animal and then determining which somatic cell hybrid(s) expresses the polypeptide and noting the chromosome(s) from the first species of animal that it contains. For examples of this technique, see Pajunen *et al.* (1988) *Cytogenet. Cell Genet. 47*:37-41 and Van Keuren *et al.* (1986) *Hum. Genet. 74*:34-40. Alternatively, the presence of the polypeptide in the somatic cell hybrids can be determined by assaying an activity or property of the polypeptide, for example, enzymatic activity, as described in Bordelon-Riser *et al.* (1979) *Somatic Cell Genetics 5*:597-613 and Owerbach *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*:5640-5644.

2. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags"

which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, 27, 36, 44, 52, 58, 63, 73, 93, 110, 123 or 128 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:1, 27, 36, 44, 52, 58, 63, 73, 93, 110, 123 or 128 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from the nucleic acid sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Gene Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA

sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the nucleic acid sequences of the invention or portions thereof, *e.g.*, fragments derived from noncoding regions having a length of at least 20 or 30 bases.

The nucleic acid sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or by organ type.

C. <u>Predictive Medicine</u>

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining expression of a polypeptide or nucleic acid of the invention and/or activity of a polypeptide of the invention, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant expression or activity of a polypeptide of the invention. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, mutations in a gene of the invention can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an

individual prior to the onset of a disorder characterized by or associated with aberrant expression or activity of a polypeptide of the invention.

Another aspect of the invention provides methods for expression of a nucleic acid or polypeptide of the invention or activity of a polypeptide of the invention in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent).

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs or other compounds) on the expression or activity of a polypeptide of the invention in clinical trials. These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a polypeptide or nucleic acid (*e.g.*, mRNA, genomic DNA) of the invention such that the presence of a polypeptide or nucleic acid of the invention is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA encoding a polypeptide of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a full-length cDNA, such as the nucleic acid of SEQ ID NO:1, 27, 36, 44, 52, 58, 63, 73, 93, 110, 123 or 128, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a polypeptide of the invention. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide of the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another

reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of a polypeptide of the invention include introducing into a subject a labeled antibody directed against the polypeptide. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a polypeptide of the invention or mRNA or genomic DNA encoding a polypeptide of the invention, such that the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide is detected in the biological sample, and comparing the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the control sample with the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of a polypeptide of the invention as discussed, for example, in sections above relating to uses of the sequences of the invention.

For example, kits can be used to determine if a subject is suffering from or is at increased risk of disorders such as immunological disorders, neurological disorders, eye disorders and embryonic disorders, which are associated with aberrant TANGO 339 expression. In another example, kits can be used to determine if a subject is suffering

from or is at increased risk of disorders such as immunological disorders, *e.g.*, autoimmune disorders (*e.g.*, arthritis, graft rejection (*e.g.*, allograft rejection), T cell disorders (*e.g.*, AIDS)) and inflammatory disorders (*e.g.*, bacterial infection, psoriasis, septicemia, cerebral malaria, inflammatory bowel disease, arthritis (*e.g.*, rheumatoid arthritis, osteoarthritis), and allergic inflammatory disorders (*e.g.*, asthma, psoriasis)). Disorders associated with TANGO 339 activity also include apoptotic disorders (*e.g.*, rheumatoid arthritis, systemic lupus erythematosus, insulin-dependent diabetes mellitus), cytotoxic disorders, septic shock, cachexia, and proliferative disorders (*e.g.*, B cell cancers stimulated by TNF), which are associated with aberrant TANGO 353 expression. In another example, kits can be used to determine if a subject is suffering from or is at increased risk of disorders such as immunological disorders (*e.g.*, thymic disorders) and embryonic disorders, which are associated with aberrant TANGO 358 expression. In another example, kits can be used to determine if a subject is suffering from or is at increased risk of disorders such as immunological disorders, *e.g.*, autoimmune disorders (*e.g.*, arthritis, graft rejection (*e.g.*, allograft rejection), T cell disorders (*e.g.*, AIDS)) and inflammatory disorders (*e.g.*, bacterial infection, psoriasis, septicemia, cerebral malaria, inflammatory bowel disease, arthritis (*e.g.*, rheumatoid arthritis, osteoarthritis), and allergic inflammatory disorders (*e.g.*, asthma, psoriasis)). Disorders associated with decreased [x] activity also include apoptotic disorders (*e.g.*, rheumatoid arthritis, systemic lupus erythematosus, insulin-dependent diabetes mellitus), cytotoxic disorders, septic shock, cachexia, and proliferative disorders (*e.g.*, B cell cancers stimulated by TNF), which are associated with aberrant TANGO 368 or TANGO 369 expression. In another example, kits can be used to determine if a subject is suffering from or is at increased risk of disorders such as immunological disorders (*e.g.*, platelet disorders) endothelial disorders and embryonic disorders, which are associated with aberrant TANGO 402 expression. In another example, kits can be used to determine if a subject is suffering from or is at risk for brain-related disorders such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, *e.g.*, infection, toxins, or drugs), inflammations (*e.g.*, bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (*e.g.*, hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (*e.g.*, neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain, which are associated with aberrant MANGO 346 or MANGO 349 expression. In still another example, kits can be used to determine if a subject is suffering from or is at risk for prostate-related disorders, (*e.g.* prostate cancer, prostatitis, benign

prostatic hypertrophy, benign prostatic hyperplasia and atypical prostatic stromal lesions) which can be associated with aberrant TANGO 365 or TANGO 383 expression. In another example, kits can be used to determine if a subject is suffering from or is at risk for endocrine-related disorders, *e.g.*, whole animal homeostasis, appetite-related disorders, which are associated with aberrant TANGO 393 expression. The kit, for example, may comprise a labeled compound or agent capable of detecting the polypeptide or mRNA encoding the polypeptide in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (*e.g.*, an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits may also include instructions for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level.

For antibody-based kits, the kit may comprise, for example: (1) a first antibody (*e.g.*, attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit may comprise, for example: (1) an oligonucleotide, *e.g.*, a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule encoding a polypeptide of the invention. The kit may also comprise, *e.g.*, a buffering agent, a preservative, or a protein stabilizing agent. The kit may also comprise components necessary for detecting the detectable agent (*e.g.*, an enzyme or a substrate). The kit may also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide.

2. Prognostic Assays

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a

disorder associated with aberrant expression or activity of a polypeptide of the invention, such as a proliferative disorder, *e.g.*, psoriasis or cancer, or an angiogenic disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing such a disease or disorder. Thus, the present invention provides a method in which a test sample is obtained from a subject and a polypeptide or nucleic acid (*e.g.*, mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the polypeptide. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

The prognostic assays described herein, for example, can be used to identify a subject having or at risk of developing disorders such as disorders discussed, for example, in sections above relating to uses of the sequences of the invention. For example, prognostic assays described herein can be used to identify a subject having or at risk of developing immunological disorders, neurological disorders and embryonic disorders, which are associated with aberrant TANGO 339 expression. In another example, prognostic assays described herein can be used to identify a subject having or at risk of developing immunological disorders *e.g.*, autoimmune disorders (*e.g.*, arthritis, graft rejection (*e.g.*, allograft rejection), T cell disorders (*e.g.*, AIDS)) and inflammatory disorders (*e.g.*, bacterial infection, psoriasis, septicemia, cerebral malaria, inflammatory bowel disease, arthritis (*e.g.*, rheumatoid arthritis, osteoarthritis), and allergic inflammatory disorders (*e.g.*, asthma, psoriasis)). Disorders associated with TANGO 339 activity also include apoptotic disorders (*e.g.*, rheumatoid arthritis, systemic lupus erythematosus, insulin-dependent diabetes mellitus), cytotoxic disorders, septic shock, cachexia, and proliferative disorders (*e.g.*, B cell cancers stimulated by TNF), which are associated with aberrant TANGO 353 expression. In another example, prognostic assays described herein can be used to identify a subject having or at risk of developing immunological disorders (*e.g.*, thymic disorders) and embryonic disorders, which are associated with aberrant TANGO 358 expression. In another example, prognostic assays described herein can be used to identify a subject having or at risk of developing immunological disorders *e.g.*, autoimmune disorders (*e.g.*, arthritis, graft rejection (*e.g.*, allograft rejection), T cell disorders (*e.g.*, AIDS)) and inflammatory disorders (*e.g.*, bacterial infection, psoriasis, septicemia, cerebral malaria, inflammatory bowel disease, arthritis (*e.g.*, rheumatoid arthritis, osteoarthritis), and allergic inflammatory disorders (*e.g.*, asthma, psoriasis)). Disorders associated with TANGO 358 activity also include

apoptotic disorders (*e.g.*, rheumatoid arthritis, systemic lupus erythematosus, insulin-dependent diabetes mellitus), cytotoxic disorders, septic shock, cachexia, and proliferative disorders (*e.g.*, B cell cancers stimulated by TNF), which are associated with aberrant TANGO 368 or TANGO 369 expression. In another example, prognostic assays described herein can be used to identify a subject having or at risk of developing immunological disorders (*e.g.*, platelet disorders), endothelial disorders and embryonic disorders, which are associated with aberrant TANGO 402 expression. In another example, prognostic assays described herein can be used to identify a subject having or at risk of developing brain-related disorders such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, *e.g.*, infection, toxins, or drugs), inflammations (*e.g.*, bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (*e.g.*, hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (*e.g.*, neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain, which are associated with aberrant MANGO 346 or MANGO 349 expression. In another example, prognostic assays described herein can be used to identify a subject having or at risk of developing prostate-related disorders (*e.g*, prostate cancer, prostatis, benign prostatic hypertrophy, benign prostatic hyperplasmia and atypical prostatic stromal lesions). In another example, prognostic assays described herein can be used to identify a subject having or at risk of developing endocrine-related disorders (*e.g.*, animal homeostasis and appetite-related disorders), which are associated with aberrant TANGO 393 expression.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (*e.g.*, agents of a type which decrease activity of the polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a polypeptide of the invention in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (*e.g.*, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized aberrant expression or activity of a polypeptide of the invention. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding the polypeptide of the invention, or the mis-expression of the gene encoding the polypeptide of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from the gene; 2) an addition of one or more nucleotides to the gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; 8) a non-wild type level of a the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post-translational modification of the protein encoded by the gene. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a gene.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in a gene (*see, e.g.*, Abravaya et al. (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177),

Q-Beta Replicase (Lizardi et al. (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a selected gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al. (1996) *Human Mutation* 7:244-255; Kozal et al. (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the selected gene and detect mutations by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Bio/Techniques* 19:448), including sequencing by mass spectrometry (*see, e.g.*, PCT Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in a selected gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the technique of "mismatch cleavage" entails providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions.

In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a selected sequence, *e.g.*, a wild-type sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2766; *see also* Cotton (1993) *Mutat. Res.* 285:125-144; Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility

enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet.* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a `GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect

the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue, *e.g.*, preferably peripheral blood leukocytes, in which the polypeptide of the invention is expressed may be utilized in the prognostic assays described herein.

3. Pharmacogenomics

Agents, or modulators which have a stimulatory or inhibitory effect on activity or expression of a polypeptide of the invention as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of a polypeptide of the invention, expression of a nucleic acid of the invention, or mutation content of a gene of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. *See, e.g.*, Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main

clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of a polypeptide of the invention, expression of a nucleic acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

4. <u>Monitoring of Effects During Clinical Trials</u>

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of a polypeptide of the invention (*e.g.*, the ability to modulate aberrant cell proliferation chemotaxis, and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase gene expression, protein

levels or protein activity, can be monitored in clinical trials of subjects exhibiting decreased gene expression, protein levels, or protein activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or protein activity. In such clinical trials, expression or activity of a polypeptide of the invention and preferably, that of other polypeptide that have been implicated in for example, a cellular proliferation disorder, can be used as a marker of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including those of the invention, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates activity or expression of a polypeptide of the invention (*e.g.*, as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of a gene of the invention and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of a gene of the invention or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or nucleic acid of the invention in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level the of the polypeptide or nucleic acid of the invention in the post-administration samples; (v) comparing the level of the polypeptide or nucleic acid of the invention in the pre-administration sample with the level of the polypeptide or nucleic acid of the invention in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of the polypeptide to higher levels than detected, *i.e.*, to increase the effectiveness of the agent.

Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of the polypeptide to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

C. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, disorders characterized by aberrant expression or activity of the polypeptides of the invention include proliferative disorders such as cancer.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant expression or activity of a polypeptide of the invention, by administering to the subject an agent which modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease which is caused or contributed to by aberrant expression or activity of a polypeptide of the invention can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject. For example, an antagonist of a TANGO 339 protein may be used to treat an immunological disorder. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of a polypeptide of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. Examples of such stimulatory agents include the active polypeptide of the invention and a nucleic acid molecule encoding the polypeptide of the invention that has

been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of the polypeptide of the invention. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a polypeptide of the invention. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) expression or activity. In another embodiment, the method involves administering a polypeptide of the invention or a nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the polypeptide.

Stimulation of activity is desirable in situations in which activity or expression is abnormally low or downregulated and/or in which increased activity is likely to have a beneficial effect. Conversely, inhibition of activity is desirable in situations in which activity or expression is abnormally high or upregulated and/or in which decreased activity is likely to have a beneficial effect.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

<u>Deposit of Clones</u>

Clones containing cDNA molecules encoding TANGO 339, TANGO 353, and TANGO 358 (clones EpT339, EpT353, and EpT358, respectively), were deposited with the American Type Culture Collection (Manassas, VA) on June 29, 1999 as Accession Number PTA-292, as part of a composite deposit representing a mixture of three strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (*e.g.*, LB plates) supplemented with 100μg/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation can be digested with a combination of the restriction enzymes *Sal* I and *Not* I and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

TANGO 339 (EpT339): 2.7 kb
TANGO 353 (EpT353): 1.3 kb
TANGO 358 (EpT358): 1.6 kb

The identity of the strains can be inferred from the fragments liberated.

Clones containing cDNA molecules encoding MANGO 346, TANGO 365, and TANGO 368 (clones EpM346, EpT365, and EpT368, respectively), were deposited with the American Type Culture Collection (Manassas, VA) on June 29, 1999 as Accession Number PTA-291, as part of a composite deposit representing a mixture of three strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (*e.g.*, LB plates) supplemented with 100μg/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation can be digested with a combination of the restriction enzymes *Sal* I and *Not* I and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

MANGO 346 (EpM346): 1.2 kb
TANGO 365 (EpT365): 1.4 kb
TANGO 368 (EpT368): 1.0 kb

The identity of the strains can be inferred from the fragments liberated.

Clones containing cDNA molecules encoding MANGO 349, TANGO 369, TANGO 383, and TANGO 393 (clones EpM349, EpT369, EpT383, and EpT393, respectively), were deposited with the American Type Culture Collection (Manassas, VA) on June 29, 1999 as Accession Number PTA-295, as part of a composite deposit representing a mixture of four strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (*e.g.*, LB plates) supplemented with 100μg/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation can be digested with a combination of the restriction enzymes *Sal* I

and *Not* I and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

MANGO 349 (EpM349): 3.7 kb
TANGO 369 (EpT369): 1.1 kb
TANGO 383 (EpT383): 1.4 kb
TANGO 393 (EpT393): 1.8 kb

The identity of the strains can be inferred from the fragments liberated.

Clones containing cDNA molecules encoding TANGO 402 (clone EpT402), were deposited with the American Type Culture Collection (Manassas, VA) on June 29, 1999 as Accession Number PTA-294, as part of a composite deposit representing a mixture of two strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (*e.g.*, LB plates) supplemented with 100 μg/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation can be digested with a combination of the restriction enzymes *Sal* I and *Not* I and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

TANGO 402 (EpT402): 1.4 kb

The identity of the strain containing TANGO 402 can be inferred from the liberation of a fragment of the above-identified size.

All publications, patents and patent applications mentioned in this specification are herein incorporated by reference into the specification to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

International Application No: PCT/ /

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on pages __, lines ____ of the description '

A. IDENTIFICATION OF DEPOSIT '

Further deposits are identified on an additional sheet '

Name of depositary institution '

American Type Culture Collection

Address of depositary institution (including postal code and country) '

10801 University Blvd.
Manassas, VA 20110-2209
US

Date of deposit ' June 29, 1999 Accession Number ' PTA-291

B. ADDITIONAL INDICATIONS ' (leave blank if not applicable). This information is continued on a separate attached sheet

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE ' (if the indications are not all designated States)

D. SEPARATE FURNISHING OF INDICATIONS ' (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later ' (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☑ This sheet was received with the International application when filed (to be checked by the receiving Office)

(Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau '

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

International Application No: PCT/ /

Form PCT/RO/134 (cont.)

American Type Culture Collection

10801 University Blvd.
Manassas, VA 20110-2209
US

Accession No.	Date of Deposit
PTA-292	June 29, 1999
PTA-294	June 29, 1999
PTA-295	June 29, 1999

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 30% identical to the nucleotide sequence of SEQ ID NO:1, 2, 27 or 28, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 480 nucleotides of the nucleotide sequence of SEQ ID NO:1 or 2, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292, or a complement thereof;

c) a nucleic acid molecule comprising a fragment of at least 575 nucleotides of the nucleotide sequence of SEQ ID NO:27 or 28, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292, or a complement thereof;

d) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:3 or 29, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292;

e) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292, or a complement thereof, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:3, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292;

f) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:29, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292, or a complement thereof, wherein the fragment comprises at least 45 contiguous amino acids of SEQ ID NO:29, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292;

g) nucleic acid molecule comprising a fragment of at least 50 nucleotides of nucleic acids 1 to 2102 of SEQ ID NO:1, or a complement thereof;

h) a nucleic acid molecule comprising a fragment of at least 50 nucleotides of nucleic acids 1 to 634 of SEQ ID NO:27, or a complement thereof;

i) a nucleic acid molecule comprising a fragment of at least 150 nucleotides of SEQ ID NO:28, or a complement thereof;

j) a nucleic acid molecule comprising a fragment of at least 50 nucleotides of nucleotide sequence of SEQ ID NO:2, or complement thereof;

k) a nucleic acid molecule comprising a nucleotide sequence which is at least 30% identical to the nucleotide sequence of SEQ ID NO:36, 37, 44, 45, 58 or 59, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292 or Accession Number PTA-291, or a complement thereof;

l) a nucleic acid molecule comprising a nucleotide sequence which is at least 45% identical to the nucleotide sequence of SEQ ID NO:110 or 111, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-294, or a complement thereof;

m) a nucleic acid molecule comprising a fragment of at least 50 nucleotides of the nucleotide sequence of SEQ ID NO:36, 37, 44, 45, 58, 59, 110 or 111, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA- 292, Accession Number PTA-291 or Accession Number PTA- 294, or a complement thereof;

n) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:38, 46, 60, or 112, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292, Accession Number PTA-291 or Accession Number PTA-294; and

o) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:38, 46, 60 or 112, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292, Accession Number PTA-291 or Accession Number PTA-294,

wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:38, 46, 60 or 112, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-292, Accession Number PTA-291 or Accession Number PTA-295;

p) a nucleic acid molecule comprising a nucleotide sequence which is at least 98% identical to the nucleotide sequence of SEQ ID NO:52 or 53, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, or a complement thereof;

q) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:54, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291;

r) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:54, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:54, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the Accession Number PTA-291;

s) a nucleic acid molecule comprising a nucleotide sequence which is at least 30% identical to the nucleotide sequence of SEQ ID NO:73, 74, 93, 94, 123, 124, 128 OR 129, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291 or Accession Number PTA-295, or a complement thereof;

t) a nucleic acid molecule comprising a fragment of at least 450 nucleotides of the nucleotide sequence of SEQ ID NO:123, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, or a complement thereof;

u) a nucleic acid molecule comprising a fragment of at least 50 nucleotides of the nucleotide sequence of SEQ ID NO:74, 124, 128 or 129, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, Accession Number PTA-291, or a complement thereof;

v) a nucleic acid molecule comprising a fragment of at least 50 nucleotides of nucleic acids 1-1250 of SEQ ID NO:73, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, Accession Number PTA-291, or a complement thereof;

w) a nucleic acid molecule comprising a fragment of at least 250 nucleotides of the nucleotide sequence of SEQ ID NO:93, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, Accession Number PTA-291, or a complement thereof;

x) a nucleic acid molecule comprising a fragment of at least 200 nucleotides of the nucleotide sequence of SEQ ID NO:94, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, Accession Number PTA-291, or a complement thereof;

y) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:75, 95, 125 or 130, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291 or Accession Number PTA-295, or a complement thereof;

z) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 95, 125 or 130, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291 or Accession Number PTA-295, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:95, 125, or 130, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291 or Accession Number PTA-295, or a complement thereof;

aa) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 75, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291 or Accession Number PTA-295, wherein the fragment comprises at least 60 contiguous amino acids of SEQ ID NO:75, or the amino

acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291 or Accession Number PTA-295, or a complement thereof;

ab) a nucleic acid molecule comprising a nucleotide sequence which is at least 40% identical to the nucleotide sequence of SEQ ID NO:63, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, or a complement thereof;

ac) a nucleic acid molecule comprising a nucleotide sequence which is at least 65% identical to the nucleotide sequence of SEQ ID NO:64, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, or a complement thereof;

ad) a nucleic acid molecule comprising a fragment of at least 510 nucleotides of the nucleotide sequence of SEQ ID NO:63, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, or a complement thereof;

ae) a nucleic acid molecule comprising a fragment of at least 270 nucleotides of the nucleotide sequence of SEQ ID NO:64, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, or a complement thereof;

af) a nucleic acid molecule comprising a fragment of at least 20 least nucleotides of nucleic acids 1-255 of SEQ ID NO:64, or a complement thereof;

ag) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:65, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295;

ah) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:65, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295, wherein the fragment comprises at least 90 contiguous amino acids of SEQ ID NO:65, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-295;

ai) a nucleic acid molecule comprising a fragment of at least 20 nucleotides of nucleic acids 775 to 1386 of SEQ ID NO:63, or a complement thereof; and

aj) a nucleic acid molecule comprising a fragment of at least 20 nucleotides of nucleic acids 1 to 984 of SEQ ID NO:93, or a complement thereof; and

ak) a nucleic acid molecule comprising a fragment of at least 20 nucleotides of nucleic acids 1666 to 1946 of SEQ ID NO:93, or a complement thereof.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 2, 27, 28, 36, 37, 44, 45, 52, 53, 58, 59, 63, 64, 73, 74, 93, 94, 110, 111, 123, 124, 128, 129, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 or 230, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295;

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125; 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231;

b) an allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or the amino acid sequence encoded by the cDNA insert of plasmids deposited with the ATCC® as Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:2, 28, 37, 45, 53, 59, 64, 74, 94, 111, 124, or 129, or a complement thereof under stringent conditions; and

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 30% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:2, 28, 37, 53, 59, 64, 74, 94, 111, 124, or 129, or at least 98% to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:2, 28, 37, 53, 59, 64, 74, 94, 111, 124, or 129, or a complement thereof.

d) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 98% to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:45, or a complement thereof.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. The antibody of claim 11, wherein the antibody is a monoclonal antibody.

13. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295; and

c) an allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 29, 38, 46, 54, 60, 65, 75, 95, 112, 125, 130, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229 or 231, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-291, Accession Number PTA-292, Accession Number PTA-294, Accession Number PTA-295, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, 27, 36, 44, 52, 58, 63, 73, 93, 110, 123, or 128, or a complement thereof under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

14. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

15. The method of claim 14, wherein the compound which binds to the polypeptide is an antibody.

16. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

17. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

18. The method of claim 17, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

19. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

20. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

21. The method of claim 20, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;
c) detection of binding using an assay for TANGO 339, TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO 393, TANGO 402, MANGO 346, and MANGO 365 - mediated signal transduction.

22. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

23. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:
a) contacting a polypeptide of claim 8 with a test compound; and
b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

Input file T339; Qutput File T339.pat
Sequence length 2715

AACTTCCTCGGCCGAGCCGGGCCGCGCCGCCGCTGCCGCCGCCGCGCGGCGTTACATCCTTACTATTCAAAGTGCCA 79

TCCACAGACCTGCTGATGGGCAGCATGAGCACCACCTGGGAGCTTGCTGCGCTGTAGAATCTTGAGGGGTCTCCATCCA 158

M H Y Y R Y S 7
GATCAGCTGAATCAGAGTTTGCATTGTTAACAAGATTCTGCTTCTCAGAAG ATG CAC TAT TAT AGA TAC TCT 230

N A K V S C W Y K Y L L F S Y N I I F W 27
AAC GCC AAG GTC AGC TGC TGG TAC AAG TAC CTC CTT TTC AGC TAC AAC ATC ATC TTC TGG 290

L A G V V F L G V G L W A W S E K G V L 47
TTG GCT GGA GTT GTC TTC CTT GGA GTC GGG CTG TGG GCA TGG AGC GAA AAG GGT GTG CTG 350

S D L T K V T R M H G I D P V V L V L M 67
TCC GAC CTC ACC AAA GTG ACC CGG ATG CAT GGA ATC GAC CCT GTG GTG CTG GTC CTG ATG 410

V G V V M F T L G F A G C V G A L R E N 87
GTG GGC GTG GTG ATG TTC ACC CTG GGG TTC GCC GGC TGC GTG GGG GCT CTG CGG GAG AAT 470

I C L L N F F C G T I V L I F F L E L A 107
ATC TGC TTG CTC AAC TTT TTC TGT GGC ACC ATC GTG CTC ATC TTC TTC CTG GAG CTG GCT 530

V A V L A F L F Q D W V R D R F R E F F 127
GTG GCC GTG CTG GCC TTC CTG TTC CAG GAC TGG GTG AGG GAC CGG TTC CGG GAG TTC TTC 590

E S N I K S Y R D D I D L Q N L I D S L 147
GAG AGC AAC ATC AAG TCC TAC CGG GAC GAT ATC GAT CTG CAA AAC CTC ATC GAC TCC CTT 650

Q K A N Q C C G A Y G P E D W D L N V Y 167
CAG AAA GCT AAC CAG TGC TGT GGC GCA TAT GGC CCT GAA GAC TGG GAC CTC AAC GTC TAC 710

F N C S G A S Y S R E K C G V P F S C C 187
TTC AAT TGC AGC GGT GCC AGC TAC AGC CGA GAG AAG TGC GGG GTC CCC TTC TCC TGC TGC 770

V P D P A Q K V V N T Q C G Y D V R I Q 207
GTG CCA GAT CCT GCG CAA AAA GTT GTG AAC ACA CAG TGT GGA TAT GAT GTC AGG ATT CAG 830

L K S K W D E S I F T K G C I Q A L E S 227
CTG AAG AGC AAG TGG GAT GAG TCC ATC TTC ACG AAA GGC TGC ATC CAG GCG CTG GAA AGC 890

W L P R N I Y I V A G V F I A I S L L Q 247
TGG CTC CCG CGG AAC ATT TAC ATT GTG GCT GGC GTC TTC ATC GCC ATC TCG CTG TTG CAG 950

I F G I F L A R T L I S D I E A V K A G 267
ATA TTT GGC ATC TTC CTG GCA AGG ACG CTG ATC TCA GAC ATC GAG GCA GTG AAG GCC GGC 1010

H H F * 271
CAT CAC TTC TGA 1022

GGAGCAGAGTTGAGGGAGCCGAGCTGAGCCAGGCTGGGAGGCCAGAGCCTTTCTCTGCCATCAGCCCTACGTCCAGAGG 1101

GAGAGGAGCCGACACCCCCAGAGCCAGTGCCCCATCTTAAGCATCAGCGTGACGTGACCTCTCTGTTTCTGCTTGCTGG 1180

TGCTGAAGACCAAGGGTCCCCCTTGTTACCTGCCCAAACTTGTGACTGCATCCCTCTGGAGTCTACCCAGAGACAGAGA 1259

FIG. 1

ATGTGTCTTTATGTGGGAGTGGTGACTCTGAAAGACAGAGAGGGCTCCTGTGGCTGCCAGGAGGGCTTGACTCAGACCC 1338

CCTGCAGCTCAAGCATGTCTGCAGGACACCCTGGTCCCCTCTCCACTGGCATCCAGACATCTGCTTTGGGTCATCCACA 1417

TCTGTGGGTGGGCCGTGGGTAGAGGGACCCACAGGCGTGGACAGGGCATCTCTCTCCATCAAGCAAAGCAGCATGGGGG 1496

CCTGCCCGTAACGGGAGGCGGACGTGGCCCCGCTGGGCCTCTGAGTGCCAGCGCAGTCTGCTGGGACATGCACATATCA 1575

GGGGTTGTTTGCAGGATCCTCAGCCATGTTCAAGTGAAGTAAGCCTGAGCCAGTGCGTGGACTGGTGCCACGGGAGTGC 1654

CTTGTCCACTGTCCCCCTGTGTCCACCAGCTATTCTCCTGGCGCCGGAACTGCCTCTGGTCTTGATAGCATTAAGCCCT 1733

GATGGCGCCGGTGGCGCGGTGGGCATGGTTCTTCACTGAGAGCCGGCTCTCCTTTTCTTAAAGTGTGTAAATAGTTTAT 1812

TTATAGGGGTAAGAATGTTCTCACACCATTTCACTTCCTCTTCCTCTCCTCCAGCATTCTCCTCTGAGCAGCCTTAGAT 1891

AGTGTCCATGGCTGGAGCCGACCCTTTGAGTCCCCTTGAGTGTCTTAAGAACCAGCCCACAACAGCCTCTCTTTCTCCT 1970

CCACATACTGCAGCCTCCCTCCATGCATCCCACATACAAGCACTCCCCCACTCCCCAGCGTGGCCTCACTGTCTTCTGG 2049

TCTTGGTGCTACTGAAATTGTCACCCAGAATTTGAATCCTGACCCTCCCCACTGCAAGCCCAGGGAGCCCCAGCCCAAG 2128

ATGGCCAGCCTGAAACTGTTGGCCAGGGCTCCTCTTGTGGCCATGTACCCAGGGCTGGCTGGCCTGCCATTTGCCTCTC 2207

CCCGGAGACAGCCGTTCTTCTGCAACCACACCCCGTGCCTAGCCACAACCCCAGGCTGCAGCTGCTCAGAAGCTCCAGG 2286

CATTTTGTTTCTGGTGACCGCCCCTAATGGGATATCGGTGATCACTGGTCCACCCTTCCTGTCAGGGCTTTTCTGGGGC 2365

TGCTCTTGGAAATGAAGTCTTAAGTACTGAATAACTCCCCTGGGGATAGCTGGGGCATTTGTCTAGCTGGGCTACTTTC 2444

TAACACTTTGCCATAGCTCAGACCACTTCTCATCGTTCAGGGATGGACTGCAACCTTAATTTACTTGCCGGAGTGTACA 2523

TTCTAGTGTGGTGTATACTGGTGGCTGTTGATGATGATTTTTTTTTTTTTTTTTACACAATTCTCTGTAGACTAGGAGA 2602

AGAATGCTTGTGTTTTTCGGAAGTGTGATCCTTCTCTTTGACTGCCAAACTCTTTTATGGAATATATCTTTATATTAAT 2681

GCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2715

FIG. 1 CONTD

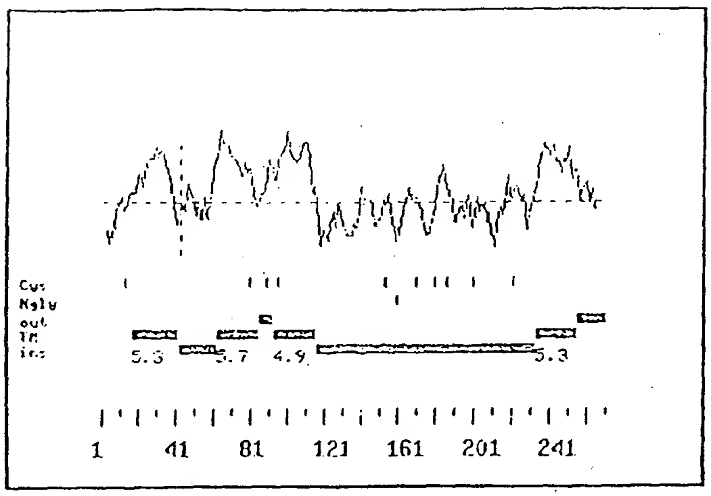

FIG. 2

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> T339 a.a.                                          270 aa vs.
> CD9 antigen a.a.                                   228 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
24.1% identity;          Global alignment score: 27

                10        20        30        40        50        60
inputs MHYYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSE---KGVLSDLTKVTRMHGIDPVVLVLM
       :      .. . :  ::::::.:.::::::.. :..::: . .   :... . :. .          : ...
       MP----VKGGTKC-IKYLLFGFNFIFWLAGIAVLAIGLWLRFDSQTKSIFEQETNNNNSSFYTGVYILIG
                   10        20        30        40        50        60

         70        80        90       100       110       120       130
inputs VGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDRFREFFESNIKSYRDD
       .:..:  .:: :: ::..:. :.:..: : ...:: .:.:.:. ..  .: :    .::. .. .  .
       AGALMMLVGFLGCCGAVQESQCMLGLFFGFLLVIFAIEIAAAIWGYSHKDEVIKEVQEFYKDTYNKLKTK
          70        80        90       100       110       120       130

       140       150       160       170       180       190       200
inputs IDLQ-NLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGVPFSCCVPDPAQKVVNTQCGYDVRI
        . : . . ... : .::: :                . . . . :              : . :..:  . :
       DEPQRETLKAIHYALNCCGLAG------------GVEQFISDIC----------PKKDVLET---FTV--
          140       150                   160                   170

         210       220       230       240       250       260       270
inputs QLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIAISLLQIFGIFLARTLISDIEAVKAGHHF
                   :.: .:.   .  .. :...: :.:... :::.... .:  .:   ..
       ------------KSCPDAIKEVFDNKFHIIGAVGIGIAVVMIFGMIFSMILCCAI--RRNREMV
                   180       190       200       210       220
```

FIG. 3

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> T339 ORF                                          810 aa vs.
> NM_001769 ORF                                     684 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
45.9% identity;          Global alignment score: 944

              10        20        30        40        50        60        70
inputs ATGCACTATTATAGATACTCTAACGCCAAGGTCAGCTGCTGGTACAAGTACCTCCTTTTCAGCTACAACA
       :::: : .: :.::...     .:.:::::      :::.   .:::.::::: :: :::.: :. :::.
       ATGC-CGGTCAAAGGAG-----GCACCAAG------TGCA---TCAAATACCTGCTGTTCGGATTTAACT
              10              20                    30        40        50

              80        90       100       110       120       130       140
inputs TCATCTTCTGGTTGGCTGGAGTTGTCTTCCTTGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGTGTGCT
       ::::::::::: : :: ::.:.:::   ::::::  .: ::.::.:::     :.. ::: .   . .  .
       TCATCTTCTGGCTTGCCGGGATTGCTGTCCTTGCCATTGGACTATGGCTCCGATTCGACTCTCAGACCAA
           60        70        80        90       100       110       120

            150       160       170       180             190       200
inputs GTCCGACCTCACCAAAGTGACCCGGATGCATGGAATCGAC--C--CTGTGGTG-CTG-GTCCTGATGGTG
       :. :..: ::.    ::::.::   . :.  ::.....  :..:   : :. .::.: ::. .: ::::::. .
       GAGCATCTTCGAGCAAGAAACTAATAATAATAATTCCAGCTTCTACACAGGAGTCTATATTCTGATCGGA
          130       140       150       160       170       180       190

          210       220       230            240       250       260       270
inputs GGCGTGGTGATGTTCACCCTGGGGTTCG-CC--GGCTGCGTGGGGGCTCTGCGGGAGAATATCTGCTTGC
       : ::  :     : .: :  ::::.:  :   ::  ::::::   :::::: :::.::::.     . :::.:::
       GCCGGCGCCCTCATGATGCTGGTGGGCTTCCTGGGCTGCTGCGGGGCTGTGCAGGAGTCCCAGTGCATGC
          200       210       220       230       240       250       260

             280       290       300       310       320       330
inputs TCAACTTTTTCTGTGGCACCATCGTGCTCATCTTCTTCCTGGAGCTGGCTGTGGCCGTGCTGG---CCTT
       : ..   : ::::  :::. : :: :: : :: :::   : : ::. :.:::: ::::.: ::::      ::
       TGGGACTGTTCTTCGGCTTCCTCTTGGTGATATTCGCCATTGAAATAGCTGCGGCCAT-CTGGGGATATT
          270       280       290       300       310       320       330

    340       350       360       370       380       390       400
inputs CCTGTTCCAGGACTGGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGAGCAACATCAAGTCCTACCGGGAC
       :: .   : ::::  .:::::  .:  .. :::.::::::: :.:.::..:: :..:::    :.: : :
       CCCA--CAAGGATGAGGTGATTAAGGAAGTCCAGGAGTTTTACAAGGACACCTACAA---CAAGCTG---
           -340       350       360       370       380            390

    410       420       430       440       450       460       470
inputs GATATCGATCTGCAAAACCTCATCGACTCCCTTCAGAAAGCTAACCAGTGCTGTGGCGCATATGGCC-CT
       .:.: :.:     : ..:.:: :: ::. .    : .:::::: :.:::     ::.::    :.:. ..:  ::
       AAAACCAAG--GATGAGCCCCAGCGGGAAACG-CTGAAAGCCATCCA---CTATG---CGTTGAACTGCT
        400       410       420       430       440                 450

    480       490       500       510       520       530       540
inputs GAAGACTGGGACCTCAACGTCTACTTCAATTGCAGCGGTGCCAGCTACAGCCGAGAGAAGTGCGGGGTCC
       :..:. :::     :: .. :   :. ..::.::   : :      : ..:..:.:::  .::::: .::    : :
       GTGGTTTGG---CTGGGGGCGTGGAACAGTTT-ATCT---CAGACATCTGCCCCAAGAAGGACG---TAC
       460       470       480       490            500       510
```

FIG. 4

```
       550       560       570       580       590       600       610
inputs CCTTCTCCTGCTGCGTGCCAGATCCTGCGCAAAAGTTGTGAACACACAGTGTGGCATATGATGTCAGGAT
        : . .::: :. ::::  :..:::::: :.    :.::    :  .::..:.::        ::    :
       TCGAAACCTTCACCGTG--AAGTCCTGTCCT----GATG----CCATCAAAGAGG--------TC----T
       520       530         540           550           560

       620       630       640       650       660       670       680
inputs TCAGCTGAAGAGCAAGTGGGATGAGTCCATCTTCACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGCTC
       ::..:...:.:              ::::             :. ::::  :::::.: .     :::.
       TCGACAATAAAT-------------TCCA-------------CATCATC--GGCGCAGTG-----GGCAT
         570                    580                      590            600

       690       700       710       720       730       740       750
inputs CCGCGGAAC-ATTTACATTGTGGCTGGCGTCTTCATCGCCATCTCGCTGTTGCAGATATTTGGCATCTTC
       : ::. ..: .:  .::: .:.  ::::.: .::.::.  :: .  ::  :::.:.      ::.  .::
       CGGCATTGCCGTGGTCATGATATTTGGCATGATCTTCAGTATGA-TCTTGTGCTGT------GCT--ATC
            610       620       630       640        650            660

        760       770       780       790       800       810
inputs CTGGCAAGGACGCTGATCTCAGACATCGAGGCAGTGAAGGCCGGCCATCACTTC
       :  :::.:.::      : :             :.::.::            ::
       C--GCAGGAAC------CGC-------------GAGATGG------------TC
             670                          680
```

FIG. 4 CONTD

```
Input file T353; Output File T353.pat
Sequence length 1239

                                                                             M      1
CTCACAGGAGGAGTTGGCGGGGAGCCTTGGGCCCCTCTGGCCTCAGCCGGATTTCCCAGCCAAACGCAGAGAGAG ATG    78

 P   W   T   I   L   L   F   A   A   G   S   L   A   I   P   A   P   S   I   R     21
CCC TGG ACC ATC TTG CTC TTT GCA GCT GGC TCC TTG GCG ATC CCA GCA CCA TCC ATC CGG   138

 L   V   P   P   Y   P   S   S   Q   E   D   P   I   H   I   A   C   M   A   P     41
CTG GTG CCC CCG TAC CCA AGC AGC CAA GAG GAC CCC ATC CAC ATC GCA TGC ATG GCC CCT   198

 G   N   F   P   G   A   N   F   T   L   Y   R   G   G   Q   V   V   Q   L   L     61
GGG AAC TTC CCG GGG GCG AAT TTC ACA CTG TAT CGA GGG GGG CAG GTG GTC CAG CTC CTG   258

 Q   A   P   T   D   Q   R   G   V   T   F   N   L   S   G   G   S   S   K   A     81
CAG GCC CCC ACG GAC CAG CGC GGG GTG ACA TTT AAC CTG AGC GGC GGC AGC AGC AAG GCT   318

 P   G   G   P   F   H   C   Q   Y   G   V   L   G   E   L   N   Q   S   Q   L    101
CCA GGG GGA CCC TTC CAC TGC CAG TAT GGA GTG TTA GGT GAG CTC AAC CAG TCC CAG CTG   378

 S   D   L   S   E   P   V   N   V   S   F   P   V   P   T   W   I   L   V   L    121
TCA GAC CTC AGC GAG CCC GTG AAC GTC TCC TTC CCA GTG CCC ACT TGG ATC TTG GTG CTC   438

 S   L   S   L   A   G   A   L   F   L   L   A   G   L   V   A   V   A   L   V    141
TCC CTG AGC CTG GCT GGT GCC CTC TTC CTC CTT GCT GGG CTG GTG GCT GTT GCC CTG GTG   498

 V   R   K   V   K   L   R   N   L   Q   K   K   R   D   R   E   S   C   W   A    161
GTC AGA AAA GTT AAA CTC AGA AAT TTA CAG AAG AAA AGA GAT CGA GAA TCC TGC TGG GCC   558

 Q   I   N   F   D   S   T   D   M   S   F   D   N   S   L   F   T   V   S   A    181
CAG ATT AAC TTC GAC AGC ACA GAC ATG TCC TTC GAT AAC TCC CTG TTT ACC GTC TCC GCG   618

 K   T   M   P   E   E   D   P   A   T   L   D   D   H   S   G   T   T   A   T    201
AAA ACG ATG CCA GAA GAA GAC CCG GCC ACC TTG GAT GAT CAC TCA GGC ACC ACT GCC ACC   678

 P   S   N   S   R   T   R   K   R   P   T   S   T   S   S   S   P   E   T   P    221
CCC AGC AAC TCC AGG ACC CGG AAG AGG CCC ACT TCC ACG TCC TCC TCG CCT GAG ACC CCC   738

 E   F   S   T   F   R   A   C   Q   *                                            231
GAA TTC AGC ACT TTC CGG GCC TGC CAG TGA                                           768

GGCTGAGGACTGGGGACCCCTCTGTCTCCAGGCATTCGGGGGCCTGAGGTCCCTCCAGCTACTTCTGGGGGGGCTCTG   847

TCAGCCACTTTCTCAGGGAATTGGACAGAGGAAAGGAAGGGGAACCCTGGCCTTGGGATTTTCATCACAGAGGAGTGGG   926

AGAGGGGACACAGGCATGGGCCTGGCACTATACAGACAACAGGAAGTTCCCCTCTCGACCTTCGGCTCCTCAGGACCAC  1005

CAGAGAAGGAGATGTCAGGACCCCTTCTTGTCCCCCAGCTCGGCCATAAGACGTCCCAGGTCTCTGCACACCCGTGGAA  1084

TTCCTCCCTTCCCCAGTGGGTTTTTGAGCATAGGGTGCCCTTGGGTGTGTTGTGTGTCTGCCTGCTGGCTTGCTTAAGT  1163

TATTAATTATAACACGGGTCAAGGTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA    1239
```

FIG. 5

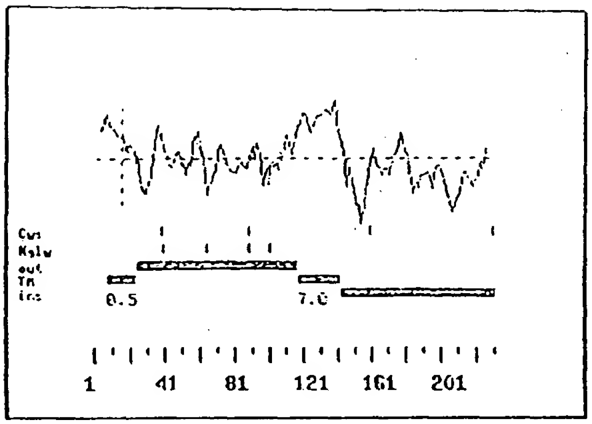

FIG. 6

```
Input file T358; Output File T358.pat
Sequence length 1608

GCACTCTAACGAAATCAAAAGTAGAAATTATGTGATACACTACAGATGGGTTTATGGGAGAGAGACAGCACAAACCCCC  79

AAACTTGGGCCTTCAGACAAAGAAAGGCTTTGGCCCTACATAAGAGTGTTGCCAACTAAATATGTATTTGCTCTAAACT  158

                            M   Y   K   L   Y   I   H   T   Y   I   C   V   Y     13
TAATAGGCTTAGGGAGGCATCGTGT ATG TAT AAA CTA TAC ATA CAT ACA TAC ATA TGT GTT TAT    222

 T   Y   T   M   P   I   M   I   L   H   L   I   F   Q   I   S   H   Q   V   L    33
ACA TAC ACA ATG CCT ATA ATG ATT CTT CAC TTA ATT TTT CAA ATT TCT CAT CAA GTA TTG  282

 V   L   I   V   P   F   K   S   A   S   V   S   I   K   S   N   L   Y   I   P    53
GTC TTA ATT GTT CCT TTT AAG AGT GCT TCT GTA AGT ATT AAA TCT AAC TTA TAT ATT CCA  342

 L   I   C   N   L   I   A   C   P   M   Y   S   S   N   N   Q   N   L   H   K    73
TTA ATT TGT AAT TTA ATT GCG TGT CCA ATG TAC AGC AGT AAC AAT CAG AAT CTT CAC AAA  402

 G   Q   C   H   F   V   K   S   F   *                                            83
GGC CAG TGC CAT TTT GTA AAA TCT TTT TAA                                          432

AGGATTAATTCAGTTATGTTATAATTAAGTATAAACATCGATATGAATACTTTCAAGTCTCCAGTTCATATTACTATAT  511

GTCAAGTGGAGAATTAATTTTTTTTTAATTTGGTAAACTAGTAGTTTAGATAGCTTCAGTGTAGGTAGTCCCTAAAAAT  590

ATCTGGGTAGACTGTGAAATTGTCTCATATGCCTTGATAATTTCAGTTTTTAAAAGGAGTAGCTTAATATTTCTATTTT  669

CTGTTCTTTTACTCAGACATTATTATTATTATTATTATTATTATTATTATTTTTTATTATTATTATTTTACAAATTGTA  748

TTTAGTATGTGTTTTTGAAACACAAAACTGGCATGCTGGTGTATCTGGTGAATTTACTATTAAGCCCAGGTCTGCACCA  827

GATGGGCAGGTGAAAATGTGATTATGAGGTGTGTGGCGTTATATCCATCTCTGTAATACTTGACTCTCCAGGTGACTTG  906

TACGTCTGTTACATTCATGAATTCACTTTACGGTTATTCTTACAGATACTGAAAGAAGCAACCTTACACCACGCAAGGT  985

CTATGGAAAACAAAAACTTCCCTCCTTTAACACAATCACAATAAATGTTTCATAACAACCAGATAAATAGAAACATGTC  1064

AGCATTTTCCCCGGCAGGTTCAGAAGTTTGATGCTGAAATACTGGTTTGCACATCTAGACCAAGACTAAAGTGAGTTTG  1143

CTATTTCAAAGGAAAATGCAAATTGAGGCTTAGGCTGCACCTTTCTGTATCACTGTTTTGGTAAGTCCACATGGGGCAG  1222

ACAAAGCCAAACACCAGCTAGGTGGTAGAATTCCCCAAATATTCATCTGTGCTGAAACCTTCAAAAGTCAGGTGCCTGG  1301

AGAGCTCATTTAATGAAAGGGTTATCTCTGCCAACCAGTCAGTAGTTGATATTTTTTTCACCAGAAGTTGTGAAATATT  1380

ATTTATGTATGTAGAAGACAGGGGCTTCCTGCCGTTGTTGTTTACTTGTGTGTTTGGGCAAGTTTCCTCCTTTCCAGAA  1459

AGCAAGTTTTCCATAAAAGCAACTTTGCAATGGCCCTACTCCTTTTGTTTGATACCCCAGGCAAAAAGCTTTTCATGGT  1538

ATTGAAATAATTTTCATTTAAAAACAGGGAGCTCAGATGAACCAGTCCTTTATTAAAAAAAAAAAAAAAA          1608
```

FIG. 7

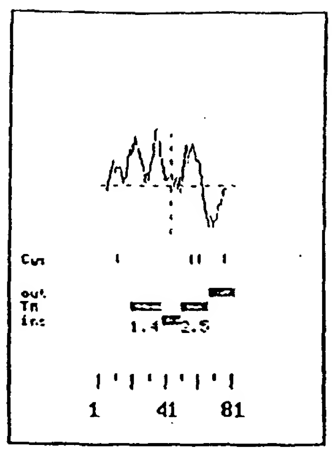

FIG. 8

Input file T365; Output File T365.pat
Sequence length 1338

```
                                                      M   L   V   A   A   L     6
GTGGATCTTCACAGTGCACGGCTTTGGGCGGCCCCTGCTGCTGTCGGCCCTGCAC ATG CTG GTG GCA GCC CTG    73

 A   C   H   R   G   A   R   R   P   M   P   G   G   T   R   C   R   V   L   L     26
GCA TGC CAC CGG GGG GCA CGG CGC CCC ATG CCA GGC GGC ACT CGC TGC CGA GTC CTA CTG   133

 L   S   L   T   F   G   T   S   M   A   C   G   N   V   G   L   R   A   V   P     46
CTC AGT CTC ACC TTT GGC ACG TCC ATG GCC TGC GGC AAC GTG GGC CTA AGG GCT GTG CCC   193

 L   D   L   A   Q   L   V   T   T   T   T   P   L   F   T   L   A   L   S   A     66
CTG GAC CTG GCA CAA CTG GTT ACT ACC ACC ACA CCT CTG TTC ACC CTG GCC CTG TCG GCG   253

 L   L   L   G   R   R   H   H   P   L   Q   L   A   A   M   G   P   L   C   L     86
CTG CTG CTG GGC CGC CGC CAC CAC CCG CTT CAG TTG GCC GCC ATG GGT CCG CTC TGC CTG   313

 G   A   A   C   S   L   A   G   E   F   R   T   P   P   T   G   C   G   F   L    106
GGG GCC GCC TGC AGC CTG GCT GGA GAG TTC CGG ACA CCC CCT ACC GGC TGT GGC TTC CTG   373

 L   A   A   T   C   L   R   G   L   K   S   V   Q   Q   N   R   V   W   L   C    126
CTC GCA GCC ACC TGC CTC CGC GGA CTC AAG TCG GTT CAG CAA AAC AGG GTC TGG CTC TGT   433

 H   P   G   C   I   G   E   I   S   A   Q   Y   S   L   R   I   L   G   S   S    146
CAC CCA GGC TGC ATT GGT GAG ATC TCA GCT CAA TAC AGC CTC CGC ATC CTG GGT TCA AGT   493

 D   S   S   A   S   A   S   Q   V   P   C   C   R   R   R   G   W   T   R   *    166
GAT TCT TCT GCC TCA GCC TCC CAA GTG CCC TGC TGC AGG AGG AGA GGC TGG ACG CGG TGA   553

CCCTGCTTTACGCCACCTCGCTGCCCAGCTTCTGCCTGCTGGCGGGTGCAGCCCTGGTGCTGGAGGCTGGCGTTGCCCC   632

ACCGCCCACTGCTGGCGACTCTCGCCTCTGGGCCTGCATCCTGCTCAGCTGCCTCCTGTCTGTTCTCTATAACCTGGCC   711

AGCTTCTCCCTGCTGGCCCTCACCTCTGCCCTCACCGTCCACGTCCTGGGCAACCTCACCGTGGTGGGCAACCTCATCC   790

TGTCCCGGCTGTTGTTTGGCAGCCGCCTCAGTGCCCTCAGCTACGTGGGCATCGCACTCACTCTTTCAGGAATGTTCCT   869

TTACCACAACTGCGAGTTCGTGGCCTCCTGGGCTGCCCGTCGGGGGCTGTGGCGGAGGGACCAGCCCAGCAAGGGTCTT   948

TGAGACCTGGGGGATCTCAGGAGCCACCTGGGATGGCCCTGGCCTGAATCCAGCCTCCGCTGTGGCCATAGAAGGAATG  1027

GAGAACAGGGCTGGGCATGGTGGCTCACGCCTATAATCCCAGCACTTCCAGAGTCCGAGGTGGGTGGATCACCTGAGGC  1106

CAGGAGTTCGAGACCAGCCTGGCTAACATGGCAAAACCTCATCTCTACTAAAAATAGAAAAATTAGCTGGGCATGGTGG  1185

CGCGTGCCTATAGTCCCAGCTACATGGGAGGCTAAGGTGGGAGGATCACTTGAGCCCTGGAGATCGAGGCTGCAGTAAG  1264

CCAAGATCGCATGCTACTGCACTCCAGCCTGGGAGACAGAGCGAGACGCTGTCTCAATTAAAAAAAAAAAAAAA       1338
```

FIG. 9

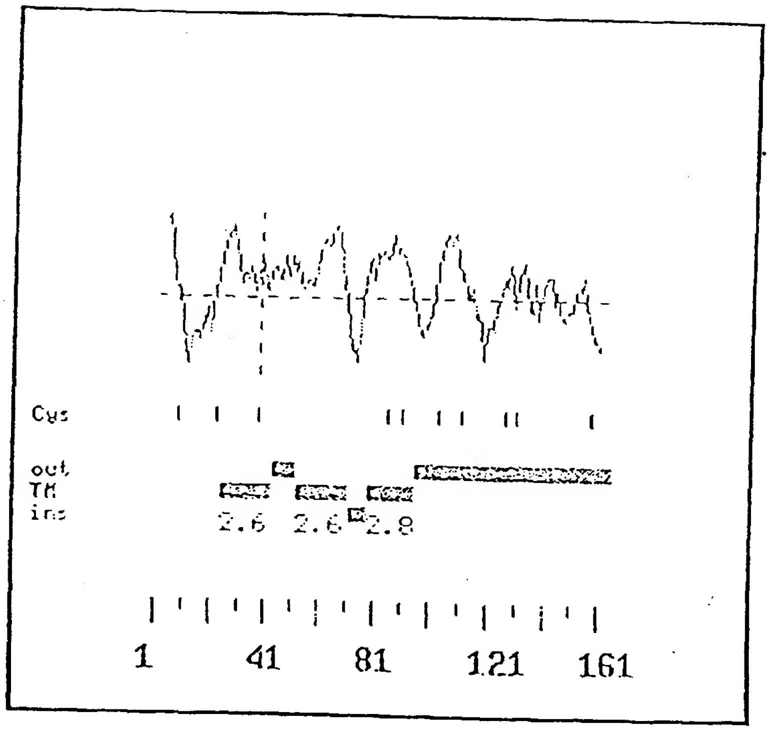

FIG. 10

Input file T368; Output File T368.pat
Sequence length 983

CGGACGCGTGGGTGTTTGATCTGCCTTTTGTGCGTGGGGTGGGAGTTAGGTAGGAATCTTAAAGTGGAGAGCCAGTTTC 79

M 1
TTCCCAAATTACTGACCTAACCCATCCTTAACCCCCAGTTCAAGGCCACCTTTGTGATAGTGAAGCTTCCAC ATG 154

L T Q P L L L S L L L Y C A C R L V L L 21
CTC ACT CAG CCC CTT CTG CTC TCT CTT CTT CTC TAC TGT GCA TGT CGG CTT GTA CTT TTG 214

P V S L K T Q P E V G W L C V H N F N F 41
CCA GTT TCT CTA AAG ACA CAA CCA GAG GTG GGG TGG CTG TGT GTG CAC AAC TTC AAC TTT 274

T C G A E S L C C I S L C K S T I C * 60
ACA TGT GGG GCT GAG TCC CTA TGT TGT ATA TCC TTG TGC AAA AGC ACA ATA TGT TAA 331

TTGCTATAGCTTTTAAAAAAATAATTAATAGTTTTTCATAATCAAATTTTCTTGCTTTTTTGTTTTTTCAAAAAAGCAT 410

ACTTTTATTGAAGAATAAACCCCTTATATATGTACACTTATTTATAACTATGAACCATGAACTAGGATAGAAATGCATT 489

GTGTATATTACAAAACATAACAAAAATAATAGGGGTAGGGAGGTGCAGATGTTGGTCAAAGGATATAAACCTGCAGTTC 568

TATGATGAATAAGTTCTGGACATCTGGAATACAGCATGGTGACTATACTTAGTAATACTATATTGTACACTTGAAGCTT 647

ACTGAAAGAGTAAATCTCAAGTGTTCTCACCACACAAACCCAAAGGTAACTATGTTCTCACCACACAAACCCAAAGGGA 726

ACTATGTATTAATTAGCTTGATTGTGGTAACCATTTCACAATGTATACATTTGCCAAAACATTATGTTGTATACCTGGA 805

ATATATAATTTTATTTATCAATTATACCTCAATAAAGCTGAAAGAGGGGATTACTAATTCCCACAAAATACAGATTTAA 884

CAAAAACTTTTATTCAACAAACAGTGCTATGAAGTTGTAAATTGGAAACAAAAGAAATAAAATTTCATCCACAGTCTTC 963

TCATCAAAAAAAAAAAAAAAA 983

FIG. 11

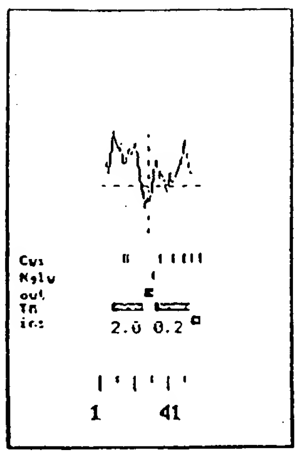

FIG. 12

```
LALIGN finds the best local alignments between two sequences
 version 2.0u54 July 1996
Please cite:
 X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

 Comparison of:
(A) inputs/nb160000.tmp > T368 n.a.                                   -
983 aa
(B) inputs/nb217750.tmp > Homo sapiens T-cell receptor gamma V1 gene regi  -
53221 aa
 using matrix file: pam120.mat, gap penalties: -12/-4

  99.3% identity in 973 aa overlap; score: 4596

                  10        20        30        40        50        60
        CGG-ACGCGTGGCTGTTTGATCTGCCTTTTGTGCGTGGGGTGGGAGTTAGGTAGGAATCTTAAAGTGGAG
        : : ::. ::.  :::::::::::::::::::::::::::::::::::::::::::::::::::::::::
        CTGCACATGTATTTGTTTGATCTGCCTTTTGTGCGTGGGGTGGGAGTTAGGTAGGAATCTTAAAGTGGAG
            46910     46920     46930     46940     46950     46960     46970

      70        80        90       100       110       120       130
        AGCCAGTTTCTTCCCAAATTACTGACCTAACCCATCCTTAACCCCCAGTTCAAGGCCACCTTTGTGATAG
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
        AGCCAGTTTCTTCCCAAATTACTGACCTAACCCATCCTTAACCCCCAGTTCAAGGCCACCTTTGTGATAG
            46980     46990     47000     47010     47020     47030     47040

     140       150       160       170       180       190       200
        TGAAGCTTCCACATGCTCACTCAGCCCCTTCTGCTCTCTCTTCTTCTCTACTGTGCATGTCGGCTTGTAC
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
        TGAAGCTTCCACATGCTCACTCAGCCCCTTCTGCTCTCTCTTCTTCTCTACTGTGCATGTCGGCTTGTAC
            47050     47060     47070     47080     47090     47100     47110

     210       220       230       240       250       260       270
        TTTTGCCAGTTTCTCTAAAGACACAACCAGAGGTGGGGTGGCTGTGTGTGCACAACTTCAACTTTACATG
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
        TTTTGCCAGTTTCTCTAAAGACACAACCAGAGGTGGGGTGGCTGTGTGTGCACAACTTCAACTTTACATG
            47120     47130     47140     47150     47160     47170     47180

     280       290       300       310       320       330       340
        TGGGGCTGAGTCCCTATGTTGTATATCCTTGTGCAAAAGCACAATATGTTAATTGCTATAGCTTTTAAAA
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
        TGGGGCTGAGTCCCTATGTTGTATATCCTTGTGCAAAAGCACAATATGTTAATTGCTATAGCTTTTAAAA
            47190     47200     47210     47220     47230     47240     47250

     350       360       370       380       390       400       410
        AAATAATTAATAGTTTTTCATAATCAAATTTTCTTGCTTTTTTGTTTTTTCAAAAAAGCATACTTTTATT
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
        AAATAATTAATAGTTTTTCATAATCAAATTTTCTTGCTTTTTTGTTTTTTCAAAAAAGCATACTTTTATT
            47260     47270     47280     47290     47300     47310     47320

     420       430       440       450       460       470       480
        GAAGAATAAACCCCTTATATATGTACACTTATTTATAACTATGAACCATGAACTAGGATAGAAATGCATT
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
        GAAGAATAAACCCCTTATATATGTACACTTATTTATAACTATGAACCATGAACTAGGATAGAAATGCATT
```

FIG. 13

```
     47330     47340     47350     47360     47370     47380     47390

490        500        510        520        530        540        550
  GTGTATATTACAAAACATAACAAAAATAATAGGGGTAGGGAGGTGCAGATGTTGGTCAAAGGATATAAAC
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
  GTGTATATTACAAAACATAACAAAAATAATAGGGGTAGGGAGGTGCAGATGTTGGTCAAAGGATATAAAC
     47400     47410     47420     47430     47440     47450     47460

560        570        580        590        600        610        620
  CTGCAGTTCTATGATGAATAAGTTCTGGACATCTGGAATACAGCATGGTGACTATACTTAGTAATACTAT
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
  CTGCAGTTCTATGATGAATAAGTTCTGGACATCTGGAATACAGCATGGTGACTATACTTAGTAATACTAT
     47470     47480     47490     47500     47510     47520     47530

630        640        650        660        670        680        690
  ATTGTACACTTGAAGCTTACTGAAAGAGTAAATCTCAAGTGTTCTCACCACACAAACCCAAAGGTAACTA
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
  ATTGTACACTTGAAGCTTACTGAAAGAGTAAATCTCAAGTGTTCTCACCACACAAACCCAAAGGTAACTA
     47540     47550     47560     47570     47580     47590     47600

700        710        720        730        740        750        760
  TGTTCTCACCACACAAACCCAAAGGGAACTATGTATTAATTAGCTTGATTGTGGTAACCATTTCACAATG
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
  TGTTCTCACCACACAAACCCAAAGGGAACTATGTATTAATTAGCTTGATTGTGGTAACCATTTCACAATG
     47610     47620     47630     47640     47650     47660     47670

770        780        790        800        810        820        830
  TATACATTTGCCAAAACATTATGTTGTATACCTGGAATATATAATTTTATTTATCAATTATACCTCAATA
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
  TATACATTTGCCAAAACATTATGTTGTATACCTGGAATATATAATTTTATTTATCAATTATACCTCAATA
     47680     47690     47700     47710     47720     47730     47740

840        850        860        870        880        890        900
  AAGCTGAAAGAGGGGATTACTAATTCCCACAAAATACAGATTTAACAAAAACTTTTATTCAACAAACAGT
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
  AAGCTGAAAGAGGGGATTACTAATTCCCACAAAATACAGATTTAACAAAAACTTTTATTCAACAAACAGT
     47750     47760     47770     47780     47790     47800     47810

910        920        930        940        950        960        970
  GCTATGAAGTTGTAAATTGGAAACAAAAGAAATAAAATTTCATCCACAGTCTTCTCATCAAAA
  :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
  GCTATGAAGTTGTAAATTGGAAACAAAAGAAATAAAATTTCATCCACAGTCTTCTCATCAAAA
     47820     47830     47840     47850     47860     47870
```

FIG. 13 CONTD

```
Input file T369; Output File T369.pat
Sequence length 1119

GGAACCCTCCTGTCCCCCTGCCCTCTCTGCTCACATCATCTGCTGCCCTAGACTTTGGAGGCTTGAATTTCTGCACGAG  79

GTGACTGCCAACAAATGACAAGAGCAGCCATTTAGTGAGCACGTAATTCATTTAATGGTGCTGAACACAGAGCAAGTGC  158

     M   S   I   L   V   R   V   H   L   Y   L   L   G   L   A   L   M   Q   S    19
TGC ATG AGT ATT TTG GTT AGG GTT CAC TTG TAC CTT TTA GGT TTA GCT CTT ATG CAA AGC  218

 L   W   F   R   S   M   C   H   P   Q   V   T   T   S   H   C   S   R   Y   G    39
CTG TGG TTC AGA TCC ATG TGT CAC CCT CAG GTC ACT ACG TCC CAC TGC AGC AGG TAT GGG  278

 E   N   H   N   H   N   T   F   P   C   S   E   F   L   S   H   I   C   L   *    59
GAG AAT CAT AAC CAT AAC ACC TTC CCT TGC AGT GAA TTT CTC TCT CAT ATT TGT CTT TAG  338

TTTGAACCACATAATAAATCTATAAGCGTATTATAGTTCCTATTCTATAGATGAGGAGACTGAGGCACACTAAGGGAAA  417

AAGTGACAAGGAAGAGACTAGAGGCTACATCTGATTTTACACCAAGTATTCATCCCACACAATGAATAGCAACCACCGG  496

ATGTTTTTCAAGATTTGAGTGAAGGCCAGAATCAAACTGAAGACTATGTGTGTGTGTGTATGTGTGTTCATATTCAAAA  575

ACACTACATACACGCTACATGTATTATATATATTATATAATATACACACATAATAACATATAACATATAAGTATATAAT  654

AGGTAATTATATATTTTTATATTAATAATATAATATATATTATATATATAATATACACACATATACATACTTTCTATAA  733

ATCTACTCCAAAAGCTTCAAGGTCTCCCAAATATCACATGACCCATAGCTAAGGCAATCCTGAAACCTGGCTGCTTGGG  812

CTAGCCTGGGGACACACGAGGAGTCAGGAACTGGGCTGTCCTTCCTTGCAATGCTCTCTCCTTCTGCCAAAACTTATCT  891

TTCTGATCTGCCTTTCCCCCAGGATCCAGGGTGGTCCTAGGAAACCCAAGGAAACGCTTCCAGCTGGAGTGCTCGGAG  970

GTGTAGGACATTGTTCTCTTCCCTTCCCGGGTCCTGTTGTTTTAGAACCTAATCAATAAAAATTAAGCTGGTAAAAAAA  1049

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  1119
```

FIG. 14

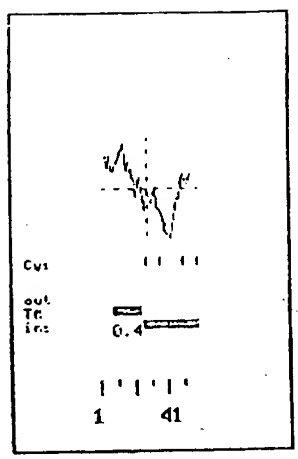

FIG. 15

```
Input file T383; Output File T383.pat
Sequence length 1386

CGGACGCGTGGGGGGGAGAATAGTCACAGCCTCCTCAAAGGGTTGTGAAAATCAAATGTGATAATTTGTGGAAAGCCCT  79

                         M   L   S   G   W   K   L   P   I   I   I   V   V       13
TAGCAGTGGCCTGGCACAAAACAA ATG CTC AGT GGA TGG AAG CTG CCT ATT ATT ATT GTC GTT     142

 V   V   V   C   H   D   C   S   G   P   G   V   E   L   A   S   G   H   V   R    33
GTT GTT GTT TGC CAT GAC TGC TCT GGG CCG GGG GTA GAG CTA GCA TCC GGG CAT GTA CGA  202

 G   K   R   E   A   G   L   Y   S   K   A   E   I   P   L   R   L   W   S   A    53
GGG AAG AGG GAG GCA GGC CTC TAT TCA AAG GCA GAA ATT CCT TTA AGA TTG TGG TCT GCT  262

 G   F   Q   G   V   S   V   L   F   V   F   V   C   L   F   V   L   R   Q   G    73
GGG TTT CAG GGA GTG TCT GTG TTG TTT GTT TTT GTT TGT TTG TTT GTT TTG AGA CAG GGT  322

 L   A   L   S   P   R   L   E   C   S   G   A   V   L   A   H   C   N   L   H    93
CTC GCT CTG TCA CCC AGG CTG GAG TGT AGT GGT GCA GTC TTG GCT CAC TGC AAC CTC CAC  382

 L   L   G   S   S   D   S   H   A   S   A   S   R   V   A   G   T   T   G   V   113
CTC CTG GGC TCA AGC GAT TCT CAT GCC TCA GCC TCC CGA GTA GCT GGG ACT ACA GGT GTG  442

 C   H   Y   A   W   L   I   F   V   F   F   V   E   T   G   F   C   H   V   A   133
TGC CAC TAT GCC TGG CTA ATT TTT GTA TTT TTT GTA GAG ACG GGG TTT TGC CAT GTT GCC  502

 Q   A   G   S   V   Y   V   *                                                   141
CAG GCT GGA AGT GTC TAT GTT TAA                                                  526

CTGCATCTTATAAACCAGCAACAAGTTTTCTACTGGGAATTAGAATGGTGCATACACAATGTATTATTATCACTGTCAG  605

ATGAGCATGCTTGAATGTAGCATGACTGCCTCTTTTTGCTTTTCCTAGAGGTTTTTTTTTTGCTTGTTACTCATCTGTT  684

GACCTACCTGGGGGAAGTAGCACCCTTGCATTTCAAAAATAAAATTGATGGCATTACAAATGGAATAGAAACCATTTTT  763

AAAATATTTTCAGTTCTCTTTCAAAATTACCTATTTTATCTTTTATATATTTGAACATATTAAGCATAGTTATTTTAAA  842

GTTCAGTTAGTCCCTTATATGGAGCCTCCATGAGTCTGTTTCTAGTGCCTGTTGTTTCTCTTGTTTTTTGTTAGTTTTG  921

TCCTGTCTCCTCCCATACTTGGCAATGTTTTAATTGAGTACAAGTAATTTCAGGTCTAGGATGGAGTTATCTTCCTCTA  1000

TGAGGATTATGTTTACATCTGGCAGGTGGCTGGGGATGCAGTGAGCCAGATCTCTTTCATCTACTTGTAGGGGATGAGA  1079

GGATTTGAGTCACTTTAAGGGCTGACCTACTTCTGTTTCATTCATATTCATAGGGTGAAGCCCACTGGGGCTCCAACCC  1158

AAAGCATGAAACATTTGCCAGGTAGCCTTCCTGTTGGTGGGCTTTGCAGTGTCTTCTAGAATCAGCAGACCCCCTACAG  1237

AAAACGTAGTCCCAGATGCCATGCATATGTCTCTGAGTTGCCTTCTTTTCCCAGATCTTGGCCTTGTAATTCTTCACTG  1316

CTTTGTTTCCTGCCCAATACTCTCAATTAGACTTAAATATATATATATACGTACAAAAAAAAAAAAAAAA          1386
```

FIG. 16

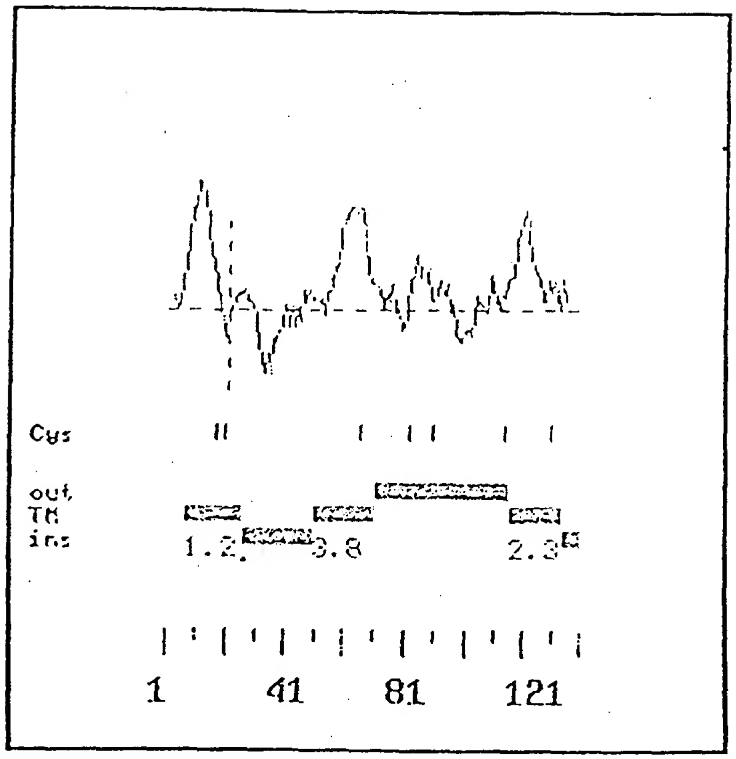

FIG. 17

```
ProDom entry 3801        Match length 73
Keywords: !!!! PROTEIN ALU SUBFAMILY WARNING ENTRY KINASE RECEPTOR ISOFORM NEURONAL

Expect 3.0e-15  Score 190        Bits 78.4       Identical 0.73  Conserved 0.78
query 79 RLECSGAVLAHCNLHLLGSSDSHASASRVAGTTG-VCHYAWLIFVF-FVETGFCHVAQAG
         RLECSGA+ AHCNL L GSSDS ASAS+VAG TG V H+A LIF F  VETGF HV QAG
sbjct  8 RLECSGAISAHCNLRLPGSSDSPASASQVAGITGDVRHHARLIFCFCLVETGFHHVGQAG


ProDom entry 85557       Match length 140
Keywords: NEURONAL THREAD PROTEIN AD7C-NTP

Expect 3.0e-03  Score 88         Bits 38.7       Identical 0.53  Conserved 0.63
query 102 ASASRVAGTTGVCHYAWLIFVF---FVETGFCHVAQAG
          ASAS+VAGT  + HY WLIF+F   F+       V QAG
sbjct 103 ASASQVAGTKDMHHYTWLIFIFIFNFLRQSLNSVTQAG
```

FIG. 18

Input file hT393; Output File hT393.pat
Sequence length 1778

M K R A S A G G S R 10
CGACTTTCAGTCCCCGACGCGCCCCGCCCAACCCCTACG ATG AAG AGG GCG TCC GCT GGA GGG AGC CGG 69

L L A W V L W L Q A W Q V A A P C P G A 30
CTG CTG GCA TGG GTG CTG TGG CTG CAG GCC TGG CAG GTG GCA GCC CCA TGC CCA GGT GCC 129

C V C Y N E P K V T T S C P Q Q G L Q A 50
TGC GTA TGC TAC AAT GAG CCC AAG GTG ACG ACA AGC TGC CCC CAG CAG GGC CTG CAG GCT 189

V P V G I P A A S Q R I F L H G N R I S 70
GTG CCC GTG GGC ATC CCT GCT GCC AGC CAG CGC ATC TTC CTG CAC GGC AAC CGC ATC TCG 249

H V P A A S F R A C R N L T I L W L H S 90
CAT GTG CCA GCT GCC AGC TTC CGT GCC TGC CGC AAC CTC ACC ATC CTG TGG CTG CAC TCG 309

N V L A R I D A A A F T G L A L L E Q L 110
AAT GTG CTG GCC CGA ATT GAT GCG GCT GCC TTC ACT GGC CTG GCC CTC CTG GAG CAG CTG 369

D L S D N A Q L R S V D P A T F H G L G 130
GAC CTC AGC GAT AAT GCA CAG CTC CGG TCT GTG GAC CCT GCC ACA TTC CAC GGC CTG GGC 429

R V H T L H L D R C G L Q E L G P G L F 150
CGC GTA CAC ACG CTG CAC CTG GAC CGC TGC GGC CTG CAG GAG CTG GGC CCG GGG CTG TTC 489

R G L A A L Q Y L Y L Q D N A L Q A L P 170
CGC GGC CTG GCT GCC CTG CAG TAC CTC TAC CTG CAG GAC AAC GCG CTG CAG GCA CTG CCT 549

D D T F R D L G N L T H L F L H G N R I 190
GAT GAC ACC TTC CGC GAC CTG GGC AAC CTC ACA CAC CTC TTC CTG CAC GGC AAC CGC ATC 609

S S V P E R A F R G L H S L D R L L L H 210
TCC AGC GTG CCC GAG CGC GCC TTC CGT GGG CTG CAC AGC CTC GAC CGT CTC CTA CTG CAC 669

Q N R V A H V H P H A F R D L G R L M T 230
CAG AAC CGC GTG GCC CAT GTG CAC CCG CAT GCC TTC CGT GAC CTT GGC CGC CTC ATG ACA 729

L Y L F A N N L S A L P T E A L A P L R 250
CTC TAT CTG TTT GCC AAC AAT CTA TCA GCG CTG CCC ACT GAG GCC CTG GCC CCC CTG CGT 789

A L Q Y L R L N D N P W V C D C R A R P 270
GCC CTG CAG TAC CTG AGG CTC AAC GAC AAC CCC TGG GTG TGT GAC TGC CGG GCA CGC CCA 849

L W A W L Q K F R G S S S E V P C S L P 290
CTC TGG GCC TGG CTG CAG AAG TTC CGC GGC TCC TCC TCC GAG GTG CCC TGC AGC CTC CCG 909

Q R L A G R D L K R L A A N D L Q G C A 310
CAA CGC CTG GCT GGC CGT GAC CTC AAA CGC CTA GCT GCC AAT GAC CTG CAG GGC TGC GCT 969

V A T G P Y H P I W T G R A T D E E P L 330
GTG GCC ACC GGC CCT TAC CAT CCC ATC TGG ACC GGC AGG GCC ACC GAT GAG GAG CCG CTG 1029

G L P K C C Q P D A A D K A S V L E P G 350
GGG CTT CCC AAG TGC TGC CAG CCA GAT GCC GCT GAC AAG GCC TCA GTA CTG GAG CCT GGA 1089

FIG. 19

```
 R   P   A   S   A   G   N   A   L   K   G   R   V   P   P   G   D   S   P   P    370
AGA CCA GCT TCG GCA GGC AAT GCG CTG AAG GGA CGC GTG CCG CCC GGT GAC AGC CCG CCG  1149

 G   N   G   S   G   P   R   H   I   N   D   S   P   F   G   T   L   P   G   S    390
GGC AAC GGC TCT GGC CCA CGG CAC ATC AAT GAC TCA CCC TTT GGG ACT CTG CCT GGC TCT  1209

 A   E   P   P   L   T   A   V   R   P   E   G   S   E   P   P   G   F   P   T    410
GCT GAG CCC CCG CTC ACT GCA GTG CGG CCC GAG GGC TCC GAG CCA CCA GGG TTC CCC ACC  1269

 S   G   P   R   R   R   P   G   C   S   R   K   N   R   T   R   S   H   C   R    430
TCG GGC CCT CGC CGG AGG CCA GGC TGT TCA CGC AAG AAC CGC ACC CGC AGC CAC TGC CGT  1329

 L   G   Q   A   G   S   G   G   G   G   T   G   D   S   E   G   S   G   A   L    450
CTG GGC CAG GCA GGC AGC GGG GGT GGC GGG ACT GGT GAC TCA GAA GGC TCA GGT GCC CTA  1389

 P   S   L   T   C   S   L   T   P   L   G   L   A   L   V   L   W   T   V   L    470
CCC AGC CTC ACC TGC AGC CTC ACC CCC CTG GGC CTG GCG CTG GTG CTG TGG ACA GTG CTT  1449

 G   P   C   *                                                                    474
GGG CCC TGC TGA                                                                  1461

CCCCCAGCGGACACAAGAGCGTGCTCAGCAGCCAGGTGTGTGTACATACGGGGTCTCTCTCCACGCCGCCAAGCCAGCC  1540

GGGCGGCCGACCCGTGGGGCAGGCCAGGCCAGGTCCTCCCTGATGGACGCCTGCCGCCCGCCACCCCCATCTCCACCCC  1619

ATCATGTTTACAGGGTTCGGCGGCAGCGTTTGTTCCAGAACGCCGCCTCCCACCCAGATCGCGGTATATAGAGATATGC  1698

ATTTTATTTTACTTGTGGAAAAATATCGGACGACGTGGAATAAAGAGCTCTTTTCTTAAAAAAAAAAAAAAAAAAAAAA  1777

A                                                                                1778
```

FIG. 19 CONTD

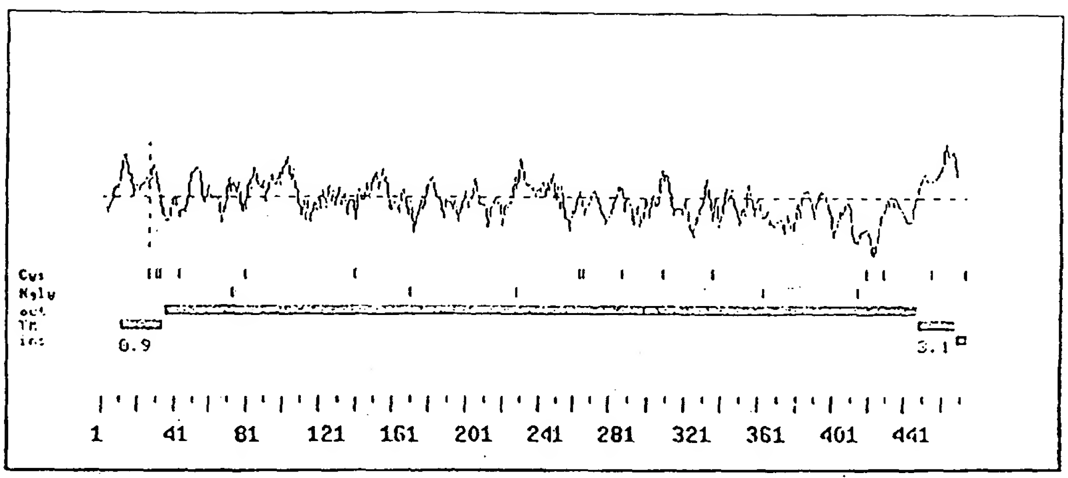

FIG. 20

```
Input file mT393; Output File mT393.pat
Sequence length 1946

CGCGCTGCGAGCGCCCCGCCAGTCCGCGCCGCCGCCCTCACCCTGTGCGCCCGCAGCCCGCGAGCCCAGCCCGGCCCGG   79

TAGAGCGGAGCGCCGGAGCCTCGTCCCGCGGCCGGGCCGGGACCGGGCCGGAGCAGCGGCGCCTGGATGCGGACCCGGC  158

                                                                        M   K   R     3
CGCGCGCAGACGGGCGCCCGCCCCGAAGCCGCTTCCAGTGCCCGACGCGCCCCGCTCGACCCCGAAG ATG AAG AGG  234

 A   S   S   G   G   S   R   L   L   A   W   V   L   W   L   Q   A   W   R   V     23
GCG TCC TCC GGA GGA AGC AGG CTG CTG GCA TGG GTG TTA TGG CTA CAG GCC TGG AGG GTA   294

 A   T   P   C   P   G   A   C   V   C   Y   N   E   P   K   V   T   T   S   C     43
GCA ACA CCA TGC CCT GGT GCT TGT GTG TGC TAC AAT GAG CCC AAG GTA ACA ACA AGC TGC   354

 P   Q   Q   G   L   Q   A   V   P   T   G   I   P   A   S   S   Q   R   I   F     63
CCC CAG CAG GGT CTG CAG GCT GTG CCC ACT GGC ATC CCA GCC TCT AGC CAG CGA ATC TTC   414

 L   H   G   N   R   I   S   H   V   P   A   A   S   F   Q   S   C   R   N   L     83
CTG CAT GGC AAC CGA ATC TCT CAC GTG CCA GCT GCG AGC TTC CAG TCA TGC CGA AAT CTC   474

 T   I   L   W   L   H   S   N   A   L   A   R   I   D   A   A   A   F   T   G    103
ACT ATC CTG TGG CTG CAC TCT AAT GCC CTG GCT CGG ATC GAT GCT GCT GCC TTC ACT GGT   534

 L   T   L   L   E   Q   L   D   L   S   D   N   A   Q   L   H   V   V   D   P    123
CTG ACC CTC CTG GAG CAA CTA GAT CTT AGT GAT AAT GCA CAG CTT CAT GTC GTG GAC CCT   594

 T   T   F   H   G   L   G   H   L   H   T   L   H   L   D   R   C   G   L   R    143
ACC ACG TTC CAC GGC CTG GGC CAC CTG CAC ACA CTG CAC CTA GAC CGA TGT GGC CTG CGG   654

 E   L   G   P   G   L   F   R   G   L   A   A   L   Q   Y   L   Y   L   Q   D    163
GAG CTG GGT CCC GGC CTA TTC CGT GGA CTA GCA GCT CTG CAG TAC CTC TAC CTA CAA GAC   714

 N   N   L   Q   A   L   P   D   N   T   F   R   D   L   G   N   L   T   H   L    183
AAC AAT CTG CAG GCA CTC CCT GAC AAC ACC TTT CGA GAC CTG GGC AAC CTC ACG CAT CTC   774

 F   L   H   G   N   R   I   P   S   V   P   E   H   A   F   R   G   L   H   S    203
TTT CTG CAT GGC AAC CGT ATC CCC AGT GTG CCT GAG CAC GCT TTC CGT GGC CTG CAC AGT   834

 L   D   R   L   L   L   H   Q   N   H   V   A   R   V   H   P   H   A   F   R    223
CTT GAC CGC CTC CTC TTG CAC CAG AAC CAT GTG GCT CGT GTG CAC CCA CAT GCC TTC CGG   894

 D   L   G   R   L   M   T   L   Y   L   F   A   N   N   L   S   M   L   P   A    243
GAC CTT GGC CGC CTC ATG ACC CTC TAC CTG TTT GCC AAC AAC CTC TCC ATG CTG CCT GCA   954

 E   V   L   M   P   L   R   S   L   Q   Y   L   R   L   N   D   N   P   W   V    263
GAG GTC CTA ATG CCC CTG AGG TCT CTG CAG TAC CTG CGA CTC AAT GAC AAC CCC TGG GTG  1014

 C   D   C   R   A   R   P   L   W   A   W   L   Q   K   F   R   G   S   S   S    283
TGT GAC TGC CGG GCA CGT CCA CTC TGG GCC TGG CTG CAG AAG TTC CGA GGT TCC TCA TCA  1074

 E   V   P   C   N   L   P   Q   R   L   A   D   R   D   L   K   R   L   A   A    303
GAG GTG CCC TGC AAC CTG CCC CAA CGC CTG GCA GAC CGT GAT CTT AAG CGC CTC GCT GCC  1134

 S   D   L   E   G   C   A   V   A   S   G   P   F   R   P   I   Q   T   S   Q    323
AGT GAC CTA GAG GGC TGT GCT GTG GCT TCA GGA CCC TTC CGT CCC ATC CAG ACC AGT CAG  1194
```

FIG. 21

```
 L   T   D   E   E   L   L   S   L   P   K   C   C   Q   P   D   A   A   D   K    343
CTC ACT GAT GAG GAG CTG CTG AGC CTC CCC AAG TGC TGC CAG CCA GAT GCT GCA GAC AAA  1254

 A   S   V   L   E   P   G   R   P   A   S   A   G   N   A   L   K   G   R   V    363
GCC TCA GTA CTG GAA CCC GGG AGG CCA GCT TCT GCC GGA AAC GCC CTC AAG GGA CGT GTG  1314

 P   P   G   D   T   P   P   G   N   G   S   G   P   R   H   I   N   D   S   P    383
CCT CCC GGT GAC ACT CCA CCA GGC AAT GGC TCA GGC CCT CGG CAC ATC AAT GAC TCT CCA  1374

 F   G   T   L   P   S   S   A   E   P   P   L   T   A   L   R   P   G   G   S    403
TTT GGA ACT TTG CCC AGC TCT GCA GAG CCC CCA CTG ACT GCC CTG CGG CCT GGG GGT TCC  1434

 E   P   P   G   L   P   T   T   G   P   R   R   R   P   G   C   S   R   K   N    423
GAG CCA CCA GGA CTT CCC ACC ACT GGT CCC CGC AGG AGG CCA GGT TGT TCC CGG AAG AAT  1494

 R   T   R   S   H   C   R   L   G   Q   A   G   S   G   A   S   G   T   G   D    443
CGC ACC CGC AGC CAC TGC CGT CTG GGC CAG GCG GGA AGT GGG GCC AGT GGA ACA GGG GAC  1554

 A   E   G   S   G   A   L   P   A   L   A   C   S   L   A   P   L   G   L   A    463
GCA GAG GGT TCA GGG GCT CTG CCT GCT CTG GCC TGC AGC CTT GCT CCT CTG GGC CTT GCA  1614

 L   V   L   W   T   V   L   G   P   C   *                                        474
CTG GTA CTT TGG ACA GTG CTT GGG CCC TGC TGA                                      1647

CCAGCCACCAGCCACCAGGTGTGTGTACATATGGGGTCTCCCTCCACGCCGCCAGCCAGAGCCAGGGACAGGCTCTGAG  1726

GGGCAGGCCAGGCCCTCCCTGACAGATGCCTCCCCACCAGCCCACCCCCATCTCCACCCCATCATGTTTACAGGGTTCC  1805

GGGGGTGGCGTTTGTTCCAGAACGCCACCTCCCACCCGGATCGCGGTATATAGAGATATGAATTTTATTTACTTGTGT   1884

AAAATATCGGATGACGTGGAATAAAGAGCTCTTTTCTTAAAAAAAAAAAAAAAAAAAAAAAA                  1946
```

FIG. 21 CONTD

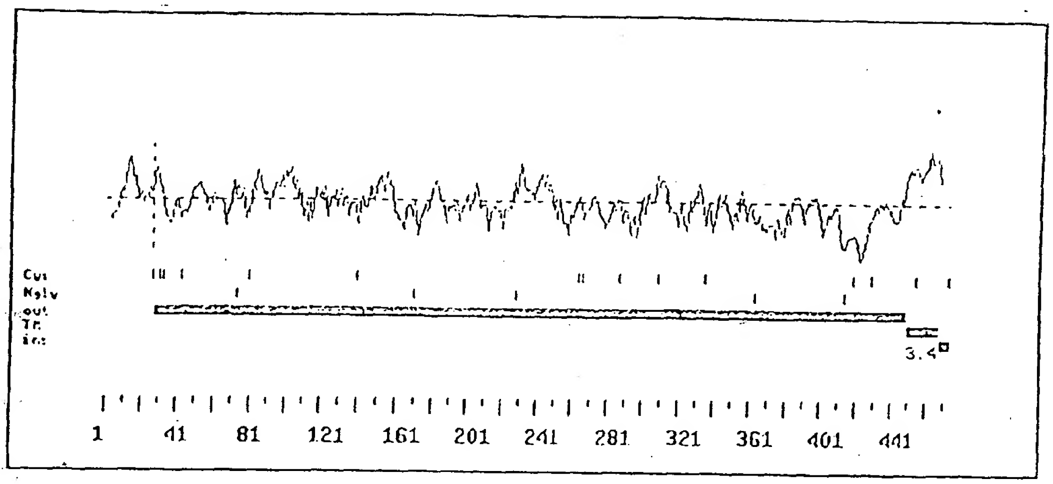

FIG. 22

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> m T393 ORF                                      1419 aa vs.
> h T393 ORF                                      1419 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
82.8% identity;          Global alignment score: 6628

                10        20        30        40        50        60        70
inputs  ATGAAGAGGGCGTCCTCCGGAGGAAGCAGGCTGCTGGCATGGGTGTTATGGCTACAGGCCTGGAGGGTAG
        ::::::::::::::: : :::::.::: :::::::::::::::::: :.:::::.:::::::::  .::::.:
        ATGAAGAGGGCGTCCGCTGGAGGGAGCCGGCTGCTGGCATGGGTGCTGTGGCTGCAGGCCTGGCAGGTGG
                10        20        30        40        50        60        70

                80        90       100       110       120       130       140
inputs  CAACACCATGCCCTGGTGCTTGTGTGTGCTACAATGAGCCCAAGGTAACAACAAGCTGCCCCCAGCAGGG
        ::.: :::::::.::::: :: ::.:::::::::::::::::::::.::.:::::::::::::::::::
        CAGCCCCATGCCCAGGTGCCTGCGTATGCTACAATGAGCCCAAGGTGACGACAAGCTGCCCCCAGCAGGG
                80        90       100       110       120       130       140

               150       160       170       180       190       200       210
inputs  TCTGCAGGCTGTGCCCACTGGCATCCCAGCCTCTAGCCAGCGAATCTTCCTGCATGGCAACCGAATCTCT
        ::::::::::::::::.  :::::::::.:: : :::::::: ::::::::::: :::::::: :::::
        CCTGCAGGCTGTGCCCGTGGGCATCCCTGCTGCCAGCCAGCGCATCTTCCTGCACGGCAACCGCATCTCG
               150       160       170       180       190       200       210

               220       230       240       250       260       270       280
inputs  CACGTGCCAGCTGCGAGCTTCCAGTCATGCCGAAATCTCACTATCCTGTGGCTGCACTCTAATGCGCTGG
        :: ::::::::::: ::::::::.    : ::::: :: ::::: ::::::::::::::::: :::: :::::
        CATGTGCCAGCTGCCAGCTTCCGTGCCTGCCGCAACCTCACCATCCTGTGGCTGCACTCGAATGTGCTGG
               220       230       240       250       260       270       280

               290       300       310       320       330       340       350
inputs  CTCGGATCGATGCTGCTGCCTTCACTGGTCTGACCCTCCTGGAGCAACTAGATCTTAGTGATAATGCACA
        : ::.:: ::::: ::::::::::::::: :::.:::::::::::::.::.:: :: :: ::::::::::
        CCCGAATTGATGCGGCTGCCTTCACTGGCCTGGCCCTCCTGGAGCAGCTGGACCTCAGCGATAATGCACA
               290       300       310       320       330       340       350

               360       370       380       390       400       410
inputs  GCTTCATGTC-GTGGACCCTACCACGTTCCACGGCCTGGGCCACCTGCACACACTGCACCTAGACCGATG
        ::: :. ::: ::::::::.::::.:::::::::::::::::::.: :.::::.::::::::.::::: ::
        GCTCCG-GTCTGTGGACCCTGCCACATTCCACGGCCTGGGCCGCGTACACACGCTGCACCTGGACCGCTG
               360       370       380       390       400       410

     420       430       440       450       460       470       480
inputs  TGGCCTGCGGGAGCTGGGTCCCGGCCTATTCCGTGGACTAGCAGCTCTGCAGTACCTCTACCTACAAGAC
        ::::::::.:::::::::: :: :: ::.::::: :: ::.::.:: :::::::::::::::::.::.:::
        CGGCCTGCAGGAGCTGGGCCCGGGGCTGTTCCGCGGCCTGGCTGCCCTGCAGTACCTCTACCTGCAGGAC
     420       430       440       450       460       470       480

     490       500       510       520       530       540       550
inputs  AACAATCTGCAGGCACTCCCTGACAACACCTTTCGAGACCTGGGCAACCTCACGCATCTCTTTCTGCATG
        :::.  ::::::::::: ::::: .::::::: :: :::::::::::::::::::.:: ::::: ::::: :
        AACGCGCTGCAGGCACTGCCTGATGACACCTTCCGCGACCTGGGCAACCTCACACACCTCTTCCTGCACG
     490       500       510       520       530       540       550

     560       570       580       590       600       610       620
inputs  GCAACCGTATCCCCAGTGT:CCTGAGCACGCTTTCCGTGGCCTGCACAGTCTTGACCGCCTCCTCTTGCA
        ::::::: ::: :::: ::::: ::::.::: :::::::: :::::::: :: ::::: :::::   ::::
        GCAACCGCATCTCCAGCGTGCCCGAGCGCGCCTTCCGTGGGCTGCACAGCCTCGACCGTCTCCTACTGCA
     560       570       580       590       600       610       620

     630       640       650       660       670       680       690
inputs  CCAGAACCATGTGGCTCGTGTGCACCCACATGCCTTCCGGGACCTTGGCCGCCTCATGACCCTCTACCTG
        :::::::::. ::::: :.:::::::::::.::::::::::: :::::::::::::::::::: ::::: :::
```

FIG. 23

```
        CCAGAACCGCGTGGCCCATGTGCACCCGCATGCCTTCCGTGACCTTGGCCGCCTCATGACACTCTATCTG
      630       640       650       660       670       680       690

      700       710       720       730       740       750       760
inputs  TTTGCCAACAACCTCTCCATGCTGCCTGCAGAGGTCCTAATGCCCCTGAGGTCTCTGCAGTACCTGCGAC
        ::::::::::: :: :: . :::::: .:.:::: :::..  :::::: :  : :::::::::::: :.:
        TTTGCCAACAATCTATCAGCGCTGCCCACTGAGGCCCTGGCCCCCCTGCGTGCCCTGCAGTACCTGAGGC
      700       710       720       730       740       750       760

      770       780       790       800       810       820       830
inputs  TCAATGACAACCCCTGGGTGTGTGACTGCCGGGCACGTCCACTCTGGGCCTGGCTGCAGAAGTTCCGAGG
        :::: :::::::::::::::::::::::::::::::: ::::::::::::::::::::::::::::: ::
        TCAACGACAACCCCTGGGTGTGTGACTGCCGGGCACGCCCACTCTGGGCCTGGCTGCAGAAGTTCCGCGG
      770       780       790       800       810       820       830

      840       850       860       870       880       890       900
inputs  TTCCTCATCAGAGGTGCCCTGCAACCTGCCCCAACGCCTGGCAGACCGTGATCTTAAGCGCCTCGCTGCC
         ::::: :: ::::::::::::::.::: :: ::::::::::::.:.:::::: :: ::.::::: :::::
        CTCCTCCTCCGAGGTGCCCTGCAGCCTCCCGCAACGCCTGGCTGGCCGTGACCTCAAACGCCTAGCTGCC
      840       850       860       870       880       890       900

      910       920       930       940       950       960       970
inputs  AGTGACCTAGAGGGCTGTGCTGTGGCTTCAGGACCCTTCCGTCCCATCCAGACCAGTCAGCTCACTGATG
        :.::::::. ::::::: :::::::: .: :: :: :.::.::::::: .:::::.:  .:  ::: ::::
        AATGACCTGCAGGGCTGCGCTGTGGCCACCGGCCCTTACCATCCCATCTGGACCGGCAGGGCCACCGATG
      910       920       930       940       950       960       970

      980       990      1000      1010      1020      1030      1040
inputs  AGGAGCTGCTGAGCCTCCCCAAGTGCTGCCAGCCAGATGCTGCAGACAAAGCCTCAGTACTGGAACCCGG
        :::::: :::::.: :: ::::::::::::::::::::::: ::.:::::::::::::::::::::.:: ::
        AGGAGCCGCTGGGGCTTCCCAAGTGCTGCCAGCCAGATGCCGCTGACAAGGCCTCAGTACTGGAGCCTGG
      980       990      1000      1010      1020      1030      1040

     1050      1060      1070      1080      1090      1100      1110
inputs  GAGGCCAGCTTCTGCCGGAAACGCCCTCAAGGGACGTGTGCCTCCCGGTGACACTCCACCAGGCAATGGC
        .::.:::::::: :: :: :: :: :: :::::::: ::::: ::::::::::  ::.::.::::: :::
        AAGACCAGCTTCGGCAGGCAATGCGCTGAAGGGACGCGTGCCGCCCGGTGACAGCCCGCCGGGCAACGGC
     1050      1060      1070      1080      1090      1100      1110

     1120      1130      1140      1150      1160      1170      1180
inputs  TCAGGCCCTCGGCACATCAATGACTCTCCATTTGGAACTTTGCCCAGCTCTGCAGAGCCCCCACTGACTG
        ::.:::::::.:::::::::::::::::::.:: ::::::.::: :::: .:::::::.:::::::::.:: ::::
        TCTGGCCCACGGCACATCAATGACTCACCCTTTGGGACTCTGCCTGGCTCTGCTGAGCCCCCGCTCACTG
     1120      1130      1140      1150      1160      1170      1180

     1190      1200      1210      1220      1230      1240      1250
inputs  CCCTGCGGCCTGGGGGTTCCGAGCCACCAGGACTTCCCACCACTGGTCCCCGCAGGAGGCCAGGTTGTTC
        :  ::::::: :.::: :::::::::::::::. : :::::::.: :: :: ::: :::::::::: :::::
        CAGTGCGGCCCCGAGGGCTCCGAGCCACCAGGGTTCCCCACCTCGGGCCCTCGCCGGAGGCCAGGCTGTTC
     1190      1200      1210      1220      1230      1240      1250

     1260      1270      1280      1290      1300      1310      1320
inputs  CCGGAAGAATCGCACCCGCAGCCACTGCCGTCTGGGCCAGGCGGGAAGTGGGGCCAGTGGAACAGGGGAC
         :: ::::: ::::::::::::::::::::::::::::::::.:: :: ::::  .: ::.::.:: :::
        ACGCAAGAACCGCACCCGCAGCCACTGCCGTCTGGGCCAGGCAGGCAGCGGGGGTGGCGGGACTGGTGAC
     1260      1270      1280      1290      1300      1310      1320

     1330      1340      1350      1360      1370      1380      1390
inputs  GCAGAGGGTTCAGGGGCTCTGCC-TGCTCTGGCCTGCAGCCTTGCTCCTCTGGGCCTTGCACTGGTACTT
         ::::.:: ::::: :: ::.:: .:: :: .:::::::::: .: :: :::::::: ::.:::::.::
        TCAGAAGGCTCAGGTGCCCTACCCAGC-CTCACCTGCAGCCTCACCCCCCTGGGCCTGGCGCTGGTGCTG
     1330      1340      1350      1360      1370      1380      1390

      1400      1410
inputs TGGACAGTGCTTGGGCCCTGC
       :::::::::::::::::::::
       TGGACAGTGCTTGGGCCCTGC
     1400      1410
```

FIG. 23 CONTD

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> hT393 a.a.                                      473 aa vs.
> mT393 a.a.                                      473 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
89.2% identity;          Global alignment score: 2279

               10        20        30        40        50        60        70
inputs MKRASAGGSRLLAWVLWLQAWQVAAPCPGACVCYNEPKVTTSCPQQGLQAVPVGIPAASQRIFLHGNRIS
       :::::.:::::::::::::::.::.:::::::::::::::::::::::::::.::::.::::::::::::
       MKRASSGGSRLLAWVLWLQAWRVATPCPGACVCYNEPKVTTSCPQQGLQAVPTGIPASSQRIFLHGNRIS
               10        20        30        40        50        60        70

               80        90       100       110       120       130       140
inputs HVPAASFRACRNLTILWLHSNVLARIDAAAFTGLALLEQLDLSDNAQLRSVDPATFHGLGRVHTLHLDRC
       :::::::..::::::::::::.::::::::::::.:::::::::::::. :::.::::::..::::::::
       HVPAASFQSCRNLTILWLHSNALARIDAAAFTGLTLLEQLDLSDNAQLHVVDPTTFHGLGHLHTLHLDRC
               80        90       100       110       120       130       140

              150       160       170       180       190       200       210
inputs GLQELGPGLFRGLAALQYLYLQDNALQALPDDTFRDLGNLTHLFLHGNRISSVPEEAFRGLHSLDRLLLH
       ::.:::::::::::::::::::::.::::::.::::::::::::::::::.::::.::::::::::::::
       GLRELGPGLFRGLAALQYLYLQDNNLQALPDNTFRDLGNLTHLFLHGNRIPSVPEHAFRGLHSLDRLLLH
              150       160       170       180       190       200       210

              220       230       240       250       260       270       280
inputs QNRVAHVHPHAFRDLGRLMTLYLFANNLSALPTEALAPLRALQYLRLNDNPWVCDCRARPLWAWLQKFRG
       ::.::.::::::::::::::::::::::: ::.:.: :::.:::::::::::::::::::::::::::::
       QNHVARVHPHAFRDLGRLMTLYLFANNLSMLPAEVLMPLRSLQYLRLNDNPWVCDCRARPLWAWLQKFRG
              220       230       240       250       260       270       280

              290       300       310       320       330       340       350
inputs SSSEVPCSLPQRLAGRDLKRLAANDLQGCAVATGPYHPIWTGRATDEEPLGLPKCCQPDAADKASVLEPG
       :::::::.::::::.::::::::.::.:::::.::..:: :.. :::: :.:::::::::::::::::::
       SSSEVPCNLPQRLADRDLKRLAASDLEGCAVASGPFRPIQTSQLTDEELLSLPKCCQPDAADKASVLEPG
              290       300       310       320       330       340       350

              360       370       380       390       400       410       420
inputs RPASAGNALKGRVPPGDSPPGNGSGPRHINDSPFGTLPGSAEPPLTAVRPEGSEPPGFPTSGPRRRPGCS
       :::::::::::::::::.::::::::::::::::::::.::::::::.:: ::::::.::.:::::::::
       RPASAGNALKGRVPPGDTPPGNGSGPRHINDSPFGTLPSSAEPPLTALRPGGSEPPGLPTTGPRRRPGCS
              360       370       380       390       400       410       420

              430       440       450       460       470
inputs RKNRTRSHCRLGQAGSGGGGTGDSEGSGALPSLTCSLTPLGLALVLWTVLGPC
       :::::::::::::::::..::::.:::::::.:.:::.:::::::::::::::
       RKNRTRSHCRLGQAGSGASGTGDAEGSGALPALACSLAPLGLALVLWTVLGPC
              430       440       450       460       470
```

FIG. 24

```
Input file T402; Output File T402.pat
Sequence length 1348

GCCAAAGAGACATATCCAAGGTTGAGATTAGTTTCCATTTTCTTTGTACTATTTTCTGGATAATAAGACATTAGACATT   79

        M   E   N   E   D   G   Y   M   T   L   S   F   K   N   R   C   K   S      18
TGAAGAG ATG GAG AAT GAA GAT GGG TAT ATG ACG CTG AGT TTC AAG AAT CGT TGT AAA TCG  140

 K   Q   K   S   K   D   F   S   L   Y   P   Q   Y   Y   C   L   L   L   I   F    38
AAG CAG AAA TCT AAA GAT TTC TCC CTA TAT CCA CAA TAT TAT TGT CTT CTG CTC ATA TTT  200

 G   C   I   V   I   L   I   F   I   M   T   G   I   D   L   K   F   W   H   K    58
GGA TGC ATT GTG ATC CTT ATA TTC ATT ATG ACA GGG ATT GAC CTG AAG TTC TGG CAT AAA  260

 K   M   D   F   S   Q   N   V   N   I   S   S   L   S   G   H   N   Y   L   C    78
AAA ATG GAT TTC TCC CAG AAT GTA AAC ATC AGC AGT CTA TCA GGA CAC AAT TAC TTG TGC  320

 P   N   D   W   L   L   N   E   G   K   C   Y   W   F   S   T   S   F   K   T    98
CCA AAT GAC TGG CTG TTG AAC GAA GGG AAA TGT TAC TGG TTT TCA ACT TCT TTT AAA ACG  380

 W   K   E   S   Q   R   D   C   T   Q   L   Q   A   H   L   L   V   I   Q   N   118
TGG AAA GAG AGT CAA CGT GAT TGT ACA CAG CTA CAG GCA CAT TTA CTG GTG ATT CAA AAT  440

 L   D   E   L   E   F   I   Q   N   S   L   K   P   G   H   F   G   W   I   G   138
TTG GAT GAG CTG GAG TTC ATA CAG AAC AGT TTA AAA CCT GGA CAT TTT GGT TGG ATT GGA  500

 L   Y   V   T   F   Q   G   N   L   W   M   W   I   D   E   H   F   L   V   P   158
CTA TAT GTT ACA TTC CAA GGG AAC CTA TGG ATG TGG ATA GAT GAA CAC TTT TTA GTT CCA  560

 E   L   F   S   V   I   G   P   T   D   D   R   S   C   A   V   I   T   G   N   178
GAA TTG TTT TCA GTG ATT GGA CCA ACT GAT GAC AGG AGC TGT GCC GTT ATC ACA GGA AAC  620

 W   V   Y   S   E   D   C   S   S   T   F   K   G   I   C   Q   R   D   A   I   198
TGG GTG TAT TCT GAA GAC TGT AGC TCC ACA TTT AAG GGC ATT TGC CAG AGA GAT GCG ATC  680

 L   T   H   N   G   T   S   G   V   *                                           208
TTG ACG CAC AAT GGA ACC AGT GGT GTG TAA                                          710

ATGTACAACCAAATATAGAAATACTTTGCATGTTAAAGCAGAGCTAGATTTTAAAGACTTAAGATTTTTAGATAAAGTT  789

TCTAACAGAAAGTTTCTGCTAACAGACATCATCTAAATAGGAGAAAAGTATTTTATCCTGAATTGACTATAAAGACAAC  868

TTCTGAACAGAACTTTTACTCTATACTTGGATTTCTGGTTTGTCTTTTCCATGGCATTGACAAGAAAAGCTAAATAAAA  947

AATTAGTAATTATTTTAATAGTTATTTAATAGTTTGATTTTTTTGCATTTAAAATAGCATAGAATAAAACAACTTTAAA  1026

GGAATGTTATTTAGCTATATGTGCTATGTGGTAGATTGGAAGGAAAGAAGCAGTATATGTACAAATATAATATTTGAAG  1105

CATGGAATTCTGAATTTTTCATCTGTGTATTATAGCCTGAAGTGTTTGGTGGGGAGTGGGTAATGAGAAATTACCTACT  1184

GGGTATAATGTACAATATTTAGGTGATGGATAAACTAAAAGCTCAGACTTCTCCACTTTGTGATATATCCATGTAACAA  1263

AATTATGCTTGTACCCTTTAAATGTATTCAAATAAAATAAAATAAAGTCATGTGGCCAAATATTCAAAACAAAAAAAAA  1342

AAAAAA                                                                           1348
```

FIG. 25

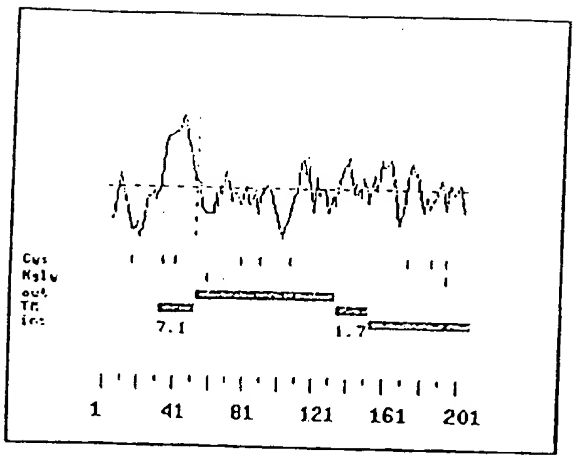

FIG. 26

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> AB010710 a.a.                                       273 aa vs.
> T402 a.a.                                           207 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
25.1% identity;          Global alignment score: -92

                   10        20        30        40        50        60
inputs MTFDD-LKIQTVKDQPDEKSNGKKAKGLQFLYSPWWCLAAATLGVLCLGLVVTIMVLGMQLSQVSDLLTQ
       :  .: .. . :..  :  .:.:.. ::.. ::   ..: :. ... ::. :..:
       MENEDGYMTLSFKNRCKSK---QKSKDFS-LYPQYYCLLL-IFG--CIVILIFIMT-GIDL---------
                  10           20         30          40        50

          70        80        90       100       110       120       130
inputs EQANLTHQKKKLEGQISARQQAEEASQESENELKEMIETLARKLNEKSKEQMELHHQNLNLQETLKRVAN
          .. :.:. .           ::. .      : .:.              :. :
       ---KFWHKKMDF-------------SQNVN------ISSLSG-------------HNYL-----------
                  60                       70

         140       150       160       170       180       190       200
inputs CSAPCPQDWIWHGENCYLFSSGSFNWEKSQEKCLSLDAKLLKINSTADLDFIQQAISYSSFPFWMGLSRR
          ::.::. .  .:: ::..  .:  ::  :  :.: :: :.. ..:.:::... . :  :.::
       ----CPNDWLLNEGKCYWFSTSFKTWKESQRDCTQLQAHLLVIQNLDELEFIQNSLKPGHFG-WIGLYVT
               80        90       100       110       120       130        140

         210       220       230       240       250       260       270
inputs NPSYPWLWEDGSPLMPHLFRVRGAVSQTYPSGTCAYIQRGAVYAENCILAAFSICQKKANL-----RAQ
        ..  :.: :  :.: :: : :....     .:: :  . ::.:.: .  .:::. : :     .
       FQGNLWMWIDEHFLVPELFSVIGPTDDR----SCAVITGNWVYSEDCSSTFKGICQRDAILTHNGTSGV
               150       160       170          180       190       200
```

FIG. 27

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> LOX-1 ORF                                      819 aa vs.
> T402 ORF                                       621 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
42.0% identity;         Global alignment score: 462

                10        20        30        40        50        60        70
inputs ATGACTTTTGATGACCTAAAGATCCAGACTGTGAAGGACCAGCCTGATGAGAAGTCAAATGGAAAAAAAG
       :::.  ..:::.::   ...:..  .:::  :::.  .: ::  : .: . ::.::.::   . :.:::
       ATGGAGAATGAAGA---TGGGTATATGACGCTGAGTTTCAAGAATCGTTGTAAATCGAA---GCAGAAAT
                10          20        30        40        50          60

                80        90       100       110       120       130       140
inputs CTAAAGGTCTTCAGTTTCTTTACTCTCCATGGTGGTGCCTGGCTGCTGCGACTCTAGGGGTCCTTTGCCT
       :::::::.: :::  :  ::.::  :.: ::. :. :: ::  ::::: :.  : :. ::.: :.:::  :
       CTAAAGAT-TTC--TCCCTATATCCACAATATTATTGTCTT-CTGCT-CA--TATTTGGATGCATTG--T
           70           80        90       100            110       120

               150       160       170       180       190       200       210
inputs GGGATTAGTAGTGACCATTATGGTGCTGGGCATGCAATTATCCCAGGTGTCTGACCTCCTAACACAAGAG
       :.    :..:: :   ::::::::.  :.:::       :::..:: ...  ::::.: :  .:: : :.:..
       GATCCTTATATT---CATTATGA--CAGGG------ATTGACCTGAAGTTCTGGCAT--AAAAAATGGA
         130         140         150            160       170        180

               220       230       240       250       260       270       280
inputs CAAGCAAACCTAACTCACCAGAAAAAGAAACTGGAGGGACAGATCTCAGCCCGGCAACAAGCAGAAGAAG
                 :..::: :::::::... ::::      ..:::    ::: :  . ::. : .::.... .
       ----------TTTCTC-CCAGAATGT-AAAC------ATCAG----CAGTCTATCAGGACACAATTACTT
                       190       200                  210       220       230

               290       300       310       320       330       340
inputs CTTCACAGGAGTCAGA-AAACGAACTCAAGGAAATGATAGAAACCCTTGCTCGGAAGCTGAATGAGAAAT
       : : ::...:.:.:. ... ::::  :.:::::::.::       :    :  ::    :.::   .: ..:::
       GTGCCCAAATGACTGGCTGTTGAACGAAGGGAAATGTTA----CTGGTTTTC---AACTTCTTTAAAA-
              240       250       260            270           280       290

   350       360       370       380       390       400       410
inputs CCAAAGAGCAAATGGAACTTCACCACCAGAATCTGAATCTCCAAGAAACACTGAAGAGAGTAGCAAATTG
       :...::  ::..:...:..:.   :   . : .:.:    :: ::.:    :::          : :..:...: .
       -CGTGGA--AAGAGAGTCAACGTGATTGTACACAG---CTACAGG---CAC-------ATTTACTGGTGA
              300       310       320         330                 340

   420       430       440       450       460       470       480
inputs TTCAGCTCCTTGTCCGCAAGACTGGATCTGGCATGGAGAAAACTGTTACCTATTTTCCTCGGGCTCATTT
       ::::. .  :::    : ....:::::  :   :::. :::  ::.:::.      ....::: ::.:  ::::
       TTCAAAAT-TTG---GATGAGCTGGAGTT--CATACAGA--ACAGTTT----AAAACCT-GGAC--ATTT
        350           360       370            380          390          400

   490       500       510       520       530       540       550
inputs AACTGGGAAAAGAGCCAAGAGAAGTGCTTGTCTTTGGATGCCAAGTTGCTGAAAATTAATAGCACAGCTG
       .. : :::...:: : :...:.   :.:.: :: ..::...::.: : : ::...::..:: : ::: ::
       TGGTTGGATTGGA-CTATATGT--TACAT-TCCAAGGGAACCTA-TGGATGTGGATAGAT-GAACA-CTT
             410       420          430       440       450        460
```

FIG. 28

```
        560       570       580       590       600       610       620
inputs ATCTGGACTTCATCCAGCAAGCAATTTCCTATTCCAGTTTTCCATTCTGGATGGGGCTGTCTCGGAGGAA
       .: .:      .:::::  :. ..::: :..:   : :  .:::. :::..   :.:.:    : :.:
       TTTAG------TTCCAG--AATTGTTTTCAGTG--ATTGGACCAA-CTGAT---GACAGGAGCTGTG---
                470       480       490       500       510

        630       640       650       660       670       680       690
inputs CCCCAGCTACCCATGGCTCTGGGAGGACGGTTCTCCTTTGATGCCCCACTTATTTAGAGTCCGAGGCGCT
         ::.  ..: :: :. .:::::.: :    ::::                   ..::.: :..::
       --CCGTTATCACAGGAAACTGGGTGTA---TTCT--------------------GAAGACTGTAGC---
          520       530       540                               550

        700       710       720       730       740       750       760
inputs GTCTCCCAGACATACCCTTCAGGTACCTGTGCATATATACAACGAGGAGCTGTTTATGCGGAAAACTGCA
          :::   ::::    :: ::: .: : :::       ::. :::..:: .: :. .::     :. ::
       ---TCC---ACAT----TTAAGG-GCATTTGC-------CAGAGAGATGCGATCTTGACG-----CA-CA
          560                570                580       590       600

        770       780       790       800       810
inputs TTTTAGCTGCCTTCAGTATATGTCAGAAGAAGGCAAACCTAAGAGCACAG
       .:  :.:      ::::. .:::                          :
       ATGGAAC------CAGTG-GTGT--------------------------G
          610                620
```

FIG. 28 CONTD

```
Input file M346; Output File M346.pat
Sequence length 1196
```

AGCATCTCTAGACCTAGAGGTTTTCTCTATTTCTCCTTTTCACTGTGACCCAGGAAATAATTTTCAGAAGTAAAAAAAT 79

CTCATCTGAGACTCTGCAACAGGCACCAGAGAGTGAGGAAGAACTTTGAGTAACAGAAACTGCTCCAATTTCTCATCCG 158

CATCTCACATCTCTGTGTCAACTATCCTTCTATCCCATTTATTCTGGTATTAGATATGTTGTCAGTGTCTCTTGTTAGG 237

TAGAGAAATCAGCAGTCAGATCTTAAGACCATTTGGTAGGTGCATCAGGAATTGACACGCAGGCCAGTTTTCCAGTCCT 316

```
    M   Y   S   F   L   C   I   L   P   L   L   L   L   A   S   C   L   L   S     19
AC ATG TAT TCT TTT CTC TGT ATC CTG CCT CTC TTG CTC TTG GCT TCC TGC CTT CTC TCC   375

 Y   S   F   L   E   Q   S   R   C   R   Q   L   E   E   L   F   P   P   S   C    39
TAC TCA TTT TTA GAA CAG TCT AGA TGC AGG CAG CTA GAA GAG TTA TTT CCT CCA AGC TGT   435

 L   G   K   G   T   I   K   E   R   F   C   T   Y   Y   D   I   K   K   E   K    59
CTA GGA AAA GGG ACA ATT AAA GAG AGA TTC TGC ACT TAT TAT GAT ATA AAA AAA GAA AAA   495

 Q   *                                                                            61
CAA TGA                                                                           501
```

GAGAACATAGTACTCTTTTACCTGTGACGTTAAATGGGAGTCACACAGGTTTAGCTATCTGTTCTAGGAGTGGATGAAC 580

AGACTATTCTCCCATGTCACTTCTTTCCTCCTGGACACCTTCAGGGAGACAGCTGGGTGAAGAATCATACTTCTGACCT 659

CTGTCAAACAGGGTCAGATGCCGCAGAGGTTCTGAGATGATAAAGGAAGTGACAGAGGGAACCTGAGGTACCACATTTC 738

TGATTTGTCATGAAAGTCTTACCTTGCTTAAGATGACTTTTTTAATGTTCCTTTCAGGGAAAATGCCAAGTGGAATAAA 817

AACCAACATCAAGTCGGCTTCCATGCATCCCTACCAGCGGTGAGTGTGGCTGGCAACCTCGACTCCCTGGTGCTCTTTG 896

CAGAGTTGGGCAGTGAAATTACCTTTTGCTCAAGGCTCACCTAGATGGGTACAATAAAAAGAACATGGGCTTTCAGCAG 975

CAGACAAATCCCACTTCCACCACTGACTAGCTGTGTGACCTTGGACAAGTGACCTAATTTTTCTGAGCCTGTTTCTCAT 1054

TTGTAAATGGTGATAATACCTACCTCATAGGGTTGTTGTGAGGATTAAAATGAGGAAATGAATGTAAAGCACTTAGTAC 1133

AGTATATGAAATAATGGGTATTCAATAAATGATAGTTTCTACAGAAAAAAAAAAAAAAAAAAA 1196

FIG. 29

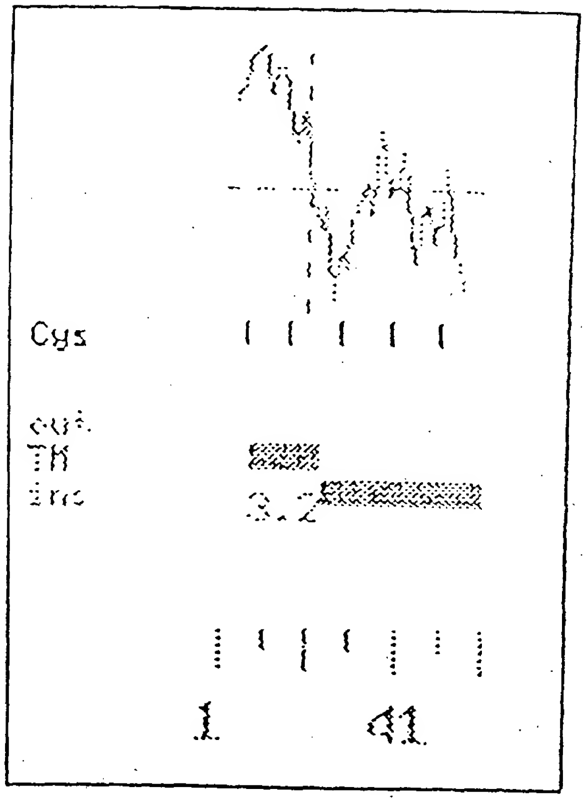

FIG. 30

```
Input file M349; Output File M349.pat
Sequence length 3649

GGCGGGCTGAGGAGCCGCCGGGCTCGGGCTCCGCGCGGGGATGTGTCTGGCCACCGCGTACTTTCGAGGGCGCCTCAAC  79

ACCCGGGGACTGATGCCCTTGGCCAGACGCACCCCCAACTGCTGTCTGCTCAGGTGTAGCCCGCGTGACTGAGTCACCG  158

                                                        M   T   A   G     4
CTGCTCCAGCTGTTTCACGTCGCCGTTCCCTTTACTGAATAGTTTGATGGGGCGCCGGGCGG ATG ACA GCG GGA  232

 T   V   V   I   T   G   G   I   L   A   T   V   I   L   L   C   I   I   A   V    24
ACG GTT GTG ATC ACT GGC GGA ATC CTA GCT ACG GTG ATC CTC CTC TGC ATC ATT GCC GTC  292

 L   C   Y   C   R   L   Q   Y   Y   C   C   K   K   S   G   T   E   V   A   D    44
CTG TGC TAC TGC AGG CTC CAG TAT TAC TGC TGC AAG AAG AGC GGA ACC GAG GTT GCA GAC  352

 E   E   E   E   R   E   H   D   L   P   T   H   P   R   G   P   T   C   N   A    64
GAG GAG GAG GAG CGG GAG CAC GAC CTT CCC ACG CAT CCC AGA GGC CCC ACC TGC AAT GCC  412

 C   S   S   Q   A   L   D   G   R   G   S   L   A   P   L   T   S   E   P   C    84
TGC AGC TCC CAA GCC CTG GAC GGC AGA GGC AGC CTG GCG CCT CTC ACC AGC GAG CCC TGC  472

 S   Q   P   C   G   V   A   A   S   H   C   T   T   C   S   P   Y   S   S   P   104
AGC CAG CCC TGT GGG GTG GCC GCG AGC CAC TGC ACT ACC TGC TCC CCA TAC AGC TCC CCC  532

 F   Y   I   R   T   A   D   M   V   P   N   G   G   G   G   E   R   L   S   F   124
TTT TAC ATA CGG ACG GCT GAC ATG GTG CCC AAT GGG GGT GGA GGC GAG AGG CTC TCC TTT  592

 A   P   T   Y   Y   K   E   G   G   P   P   S   L   K   L   A   A   P   Q   S   144
GCT CCC ACA TAC TAC AAA GAG GGG GGA CCC CCA TCC CTC AAA TTG GCA GCA CCC CAG AGT  652

 Y   P   V   T   W   P   G   S   G   R   E   A   F   T   N   P   R   A   I   S   164
TAC CCG GTG ACC TGG CCA GGC TCT GGG CGT GAG GCC TTC ACC AAT CCA AGG GCT ATT AGT  712

 T   D   V   *                                                                    168
ACA GAC GTG TAA                                                                   724

ATCCTTCCACCCCGACCCGCACACACACCCACACTGCTGCCCTGGCGGGGGCCATGGGGGTGATGAATGACCCTCCAAC  803

AGCCCCACATGGGTTGTTTCTGTTTCTTTGGCTTTTCTCGCTCCGCAGTGGAGGGTTTACTAGGATTTAAGCTTTTGAG  882

TGCATTGAGAACCAAGACAGGGCCTGGCTCCAACTCTGTGGGCCAGAGGTGGGGGACTGCTAGGTCGAGTCTGCAGCTT  961

CGCCAGTTTCTTGGTTGGGACACTCCTCTGGCAGCCCCAGCACCACCACAACCCCTTGCAGTGTGCCCCAGTCCCCTGG  1040

ATTCGCTGGACTGCAAAAAGGAGCACCAGGGAAAGCTCAGCAAAGGCTCAGGAGGCCTCCCGGGTCCTCGGGAGATGAA  1119

GCATCCGTGCCTAGGAAAAGGGGACAGAGGGCAAGGAGAGAAGTGAGAGCTACAATTCCAGCATTGGAGAAGCGAGGGG  1198

CGGGGCGTCAACGGCACTGCTTCAGGACGCGCTTGCTGAAACGACTCCAACAGCTAGTTCACAGCCCAGCTTTGTACGT  1277

TGGTTACCATAGCTACTGCTGTCACTGTAGCTGCTCCCGTAGGGACGTTGATTCCTGAACGTATCACATCTCACCTGCC  1356

CCCTTCCTCGTGGGACGTGTCAAGTTGACTTTAAAGCCTAAGGTGGCTTGTGGGGACTGCACCAGAAAGTGTCTAACCT  1435

TGTGTTCCCCGAAATCCTTTCTCTGACTTAGGAACCAGCGCCCCCTGCTGGAGAAGTTTTTTCATTTTGTTAGCCGACT  1514
```

FIG. 31

TCTGGTAGCTTAGCAAAGAGACCAGCTGACTGTCTCTCGGCCCAGCCCACCACGCCGAGTAGCTTGTGTGGATGCAGTC 1593
CTGTGAGGGTGTGCATAACCGTTCCCAGGTGTACGCACGCGCGCGCGCACACACACACACACACACACACACACACACA 1672
CACAGTAACCACGGGTAGGCACCAGGGGCTTTGTACATTCAGGAGCTGCAGAGAGGAAAGGTCTCCCATGTACCAGAGA 1751
AGGAAAAGTCTGATCCCGAAACAGCTTGAACGAAAGGTGGTTAACACGTGACCAATCCCCATGGCAGACAGCGGGCTGG 1830
TGAACAGGAAGAGCACAGAATCTCTGTGGGGCGATGCTGCAGCGGCTGAAACCACGTCAAGTCCCCCAGAGCCCGGCAT 1909
TCTATGTAAGCATCCACCTGTGACGCCGTGGAGCAGGATCTTCATTCATCTCACCAGAGCAACAGCCCAGTCCTGAGGA 1988
TGGGAGGGACGTGACTTAGCCTTGGGATTCCACGCTCGCCGCCCTCCCTTGGCCGCCCTACTCCCCAGTTTCCTGGCGA 2067
TGAGCCATTTATTGTGTCTCACACCAGCCTCATTTCCAACTCCGACTTCCTACTTAGAATGTGGCAGTGGCTTTGTGAG 2146
TGAGAATCCTGATGAGTTTTCCAGTGGCCTCTGCAGAAGCTGGCCCCCTTCTAAGGGCATAGCTAACCCCCTTCCCACT 2225
CCTGAGTCACTGAATTCGAGGAGGGTGGGGCAGGGAAAGGGCTGGCCCCCCTGCTGAAATCTTGATTCTGCATTTGAGG 2304
AAGGTCAGGGTGCCGCTGGGGGAGCAGGGAGCGGGGAGGGAGGAAGAGGGGAATGCACATCTTTTGCGCCACTATTAAG 2383
GCACCTGTCACCTTACATCTCAATCTGGAATCAAAGTGCCCTGGGTTGAGAAGCAGACCTGGGCTCTGGTCACACTTTG 2462
GTCACTCATTAGCTCTAGGACCAGTCACTAATCTCTGAGACTCCATTTTCTCCAGGGAAACGAGGCTTGCCCAGATAGA 2541
CTAACAACCCTTTAGTGCTCCAAGAAGTGAGAATTTCCGAAAAAGATCTGCCGGCCCAGAGCACTCCCTCTTGCCCTAA 2620
TCCTGGCAGGGTCTGGATGTATTTACTCAACTCTGATCTTGCGTGCAGGGATGCTTGCTGATGGATGAAAGACCACAAT 2699
AAAACAAGATTAGCAGCAAAACAATTTTAATTGTCAGTGAAAAGCAAACATGTTTGTCAGACTCTCCTGTGGAGTCATT 2778
TCAAATGGTAACTCTTCTCTTGTACCAGGAAGCAGCTGTAGGTGGGGTAATGGGATCTGTCTCCTGGTGCAGAGGGGCA 2857
GCTTATGGCGGCAGCAGCTTCAACAAGCCCGTCCCACGGCCAGCCCTAACTGGGCCAAGTGCAAGCTCCTGATGAGCTG 2936
CTGGGCCTGGCAGCGTTCCCCACTTCCTAGTGCCTTGCTCCCCTGGAATTCCACAGCCAATCAACCACAGCAGCCGTGT 3015
CCTCCTTGCCAGAAATAACATCATTCGTACGTCCTGCCTTTGTAAAAATCAAGACCCAGTTAACCATCACCCACTTCTT 3094
GCCCCTGATTGATGGGGTCCCTGAGTACAAGCATCCTGTGCAATCCCCTCCCTCCTACCGCTTTCCTAGCTAAGCCCTG 3173
CTAGCCCCACAAATCAGAATCTGGGTTTGGGGTGCTGGAGAGATGGTGGGCCCCAGCTCTTCCCCACATAGCCCACGCT 3252
GCTCAGGGCACGGTCTCCAACCAGGCTCAAGATTGCCTGCCTGCCCGCTGCAGCCGCAGGGCCCCACCCTGTGGAAAAG 3331
GAAGCTTCCCCCACTCAACTGCACTGTCCTTTTCTGCTGTCTAGTCTGTGCTAAGGGACACTATTTGTACACTTTTCAA 3410
ACGGTGCCCTAAATTGGAAGAGAAGGAGGGGCCAGGAACACAGCCCTCTCCGCCGTCTTTCTCCACCTCCACTTCACCA 3489
CTCCTGCTCTAGACTTCTTGTCTTGCATCTTTGATAACTTGGAGTCATAACCGAGTCCATGTCGGAATAAATGAGAGAT 3568
TTGAAGTNGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAA 3647
AA 3649

FIG. 31 CONTD

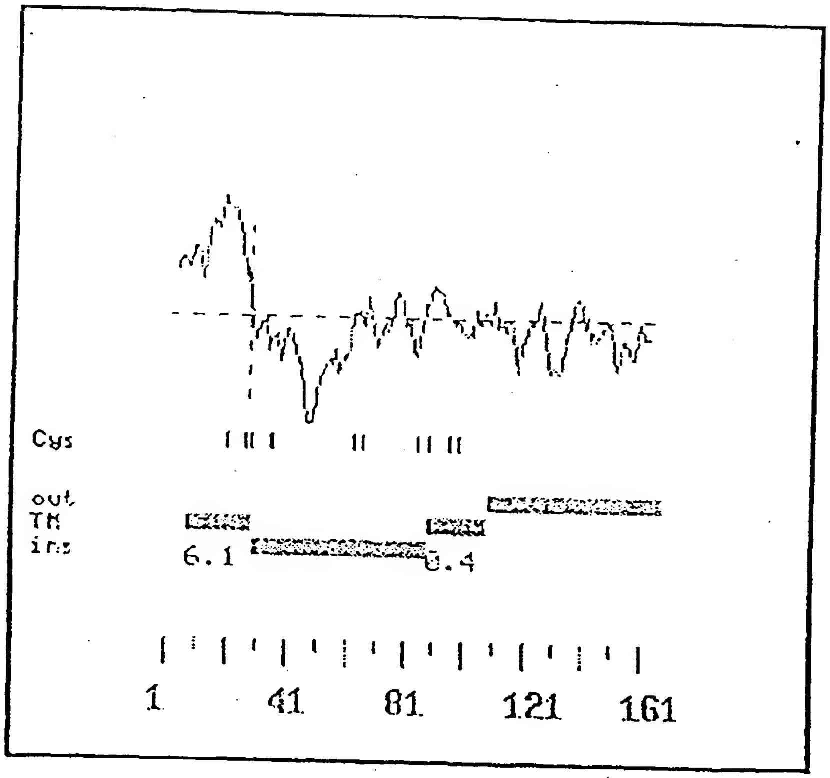

FIG. 32

SEQUENCE LISTING

<110> Millennium Pharmaceuticals, Inc.

<120> SECRETED PROTEINS AND USES THEREOF

<130> 7853-207-228

<140>
<141>

<150> 09/365,164
<151> 1999-07-30

<160> 235

<170> PatentIn Ver. 2.0

<210> 1
<211> 2715
<212> DNA
<213> Homo sapiens

<400> 1
aacttcctcg gccgagccgg gccgcgccgc cgctgccgcc gccgcgcgcg gcgttacatc 60
cttactattc aaagtgccat ccacagacct gctgatgggc agcatgagca ccacctggga 120
gcttgctgcg ctgtagaatc ttgaggggtc tccatccaga tcagctgaat cagagtttgc 180
attgttaaca agattctgct tctcagaaga tgcactatta tagatactct aacgccaagg 240
tcagctgctg gtacaagtac ctccttttca gctacaacat catcttctgg ttggctggag 300
ttgtcttcct tggagtcggg ctgtgggcat ggagcgaaaa gggtgtgctg tccgacctca 360
ccaaagtgac ccggatgcat ggaatcgacc ctgtggtgct ggtcctgatg gtgggcgtgg 420
tgatgttcac cctggggttc gccggctgcg tgggggctct gcgggagaat atctgcttgc 480
tcaacttttt ctgtggcacc atcgtgctca tcttcttcct ggagctggct gtggccgtgc 540
tggccttcct gttccaggac tgggtgaggg accggttccg ggagttcttc gagagcaaca 600
tcaagtccta ccgggacgat atcgatctgc aaaacctcat cgactccctt cagaaagcta 660
accagtgctg tggcgcatat ggccctgaag actgggacct caacgtctac ttcaattgca 720
gcggtgccag ctacagccga gagaagtgcg gggtcccctt ctcctgctgc gtgccagatc 780
ctgcgcaaaa agttgtgaac acacagtgtg gatatgatgt caggattcag ctgaagagca 840
agtgggatga gtccatcttc acgaaaggct gcatccaggc gctggaaagc tggctcccgc 900
ggaacattta cattgtggct ggcgtcttca tcgccatctc gctgttgcag atatttggca 960
tcttcctggc aaggacgctg atctcagaca tcgaggcagt gaaggccggc catcacttct 1020
gaggagcaga gttgagggag ccgagctgag ccacgctggg aggccagagc ctttctctgc 1080
catcagccct acgtccagag ggagaggagc cgacaccccc agagccagtg ccccatctta 1140
agcatcagcg tgacgtgacc tctctgtttc tgcttgctgg tgctgaagac caagggtccc 1200
ccttgttacc tgcccaaact tgtgactgca tccctctgga gtctacccag agacagagaa 1260
tgtgtcttta tgtgggagtg gtgactctga aagacagaga gggctcctgt ggctgccagg 1320
agggcttgac tcagaccccc tgcagctcaa gcatgtctgc aggacaccct ggtcccctct 1380
ccactggcat ccagacatct gctttgggtc atccacatct gtgggtgggc cgtgggtaga 1440
gggacccaca ggcgtggaca gggcatctct ctccatcaag caaagcagca tgggggcctg 1500
cccgtaacgg gaggcggacg tggccccgct gggcctctga gtgccagcgc agtctgctgg 1560
gacatgcaca tatcaggggt tgtttgcagg atcctcagcc atgttcaagt gaagtaagcc 1620
tgagccagtg cgtggactgg tgccacggga gtgccttgtc cactgtcccc ctgtgtccac 1680
cagctattct cctggcgccg gaactgcctc tggtcttgat agcattaagc cctgatggcg 1740
ccggtggcgc ggtgggcatg gttcttcact gagagccggc tctccttttc ttaaagtgtg 1800
taaatagttt atttataggg gtaagaatgt tctcacacca tttcacttcc tcttcctctc 1860
ctccagcatt ctcctctgag cagccttaga tagtgtccat ggctggagcc gaccctttga 1920
gtccccttga gtgtcttaag aaccagccca caacagcctc tctttctcct ccacatactg 1980

```
cagcctccct ccatgcatcc cacatacaag cactccccca ctccccagcg tggcctcact 2040
gtcttctggt cttggtgcta ctgaaattgt cacccagaat ttgaatcctg accctcccca 2100
ctgcaagccc agggagcccc agcccaagat ggccagcctg aaactgttgg ccagggctcc 2160
tcttgtggcc atgtacccag ggctggctgg cctgccattt gcctctcccc ggagacagcc 2220
gttcttctgc aaccacaccc cgtgcctagc cacaacccca ggctgcagct gctcagaagc 2280
tccaggcatt ttgtttctgg tgaccgcccc taatgggata tcggtgatca ctggtccacc 2340
cttcctgtca gggcttttct ggggctgctc ttggaaatga agtcttaagt actgaataac 2400
tcccctgggg atagctgggg catttgtcta gctgggctac tttctaacac tttgccatag 2460
ctcagaccac ttctcatcgt tcagggatgg actgcaacct taatttactt gccggagtgt 2520
acattctagt gtggtgtata ctggtggctg ttgatgatga tttttttttt tttttttaca 2580
caattctctg tagactagga gaagaatgct tgtgtttttc ggaagtgtga tgcttctctt 2640
tgactgccaa actcttttat ggaatatatc tttatattaa tgcaaaaaaa aaaaaaaaaa 2700
aaaaaaaaaa aaaaa                                                  2715
```

<210> 2
<211> 813
<212> DNA
<213> Homo sapiens

<400> 2

```
atgcactatt atagatactc taacgccaag gtcagctgct ggtacaagta cctccttttc 60
agctacaaca tcatcttctg gttggctgga gttgtcttcc ttggagtcgg gctgtgggca 120
tggagcgaaa agggtgtgct gtccgacctc accaaagtga cccggatgca tggaatcgac 180
cctgtggtgc tggtcctgat ggtgggcgtg gtgatgttca ccctggggtt cgccggctgc 240
gtgggggctc tgcgggagaa tatctgcttg ctcaactttt tctgtggcac catcgtgctc 300
atcttcttcc tggagctggc tgtggccgtg ctggccttcc tgttccagga ctgggtgagg 360
gaccggttcc gggagttctt cgagagcaac atcaagtcct accgggacga tatcgatctg 420
caaaacctca tcgactccct tcagaaagct aaccagtgct gtggcgcata tggccctgaa 480
gactgggacc tcaacgtcta cttcaattgc agcggtgcca gctacagccg agagaagtgc 540
ggggtcccct tctcctgctg cgtgccagat cctgcgcaaa aagttgtgaa cacacagtgt 600
ggatatgatg tcaggattca gctgaagagc aagtgggatg agtccatctt cacgaaaggc 660
tgcatccagg cgctggaaag ctggctcccg cggaacattt acattgtggc tggcgtcttc 720
atcgccatct cgctgttgca gatatttggc atcttcctgg caaggacgct gatctcagac 780
atcgaggcag tgaaggccgg ccatcacttc tga                              813
```

<210> 3
<211> 270
<212> PRT
<213> Homo sapiens

<400> 3

```
Met His Tyr Tyr Arg Tyr Ser Asn Ala Lys Val Ser Cys Trp Tyr Lys
1               5                   10                  15

Tyr Leu Leu Phe Ser Tyr Asn Ile Ile Phe Trp Leu Ala Gly Val Val
            20                  25                  30

Phe Leu Gly Val Gly Leu Trp Ala Trp Ser Glu Lys Gly Val Leu Ser
        35                  40                  45

Asp Leu Thr Lys Val Thr Arg Met His Gly Ile Asp Pro Val Val Leu
    50                  55                  60

Val Leu Met Val Gly Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys
65                  70                  75                  80

Val Gly Ala Leu Arg Glu Asn Ile Cys Leu Leu Asn Phe Phe Cys Gly
                85                  90                  95
```

Thr Ile Val Leu Ile Phe Phe Leu Glu Leu Ala Val Ala Val Leu Ala
100 105 110

Phe Leu Phe Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu
115 120 125

Ser Asn Ile Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile
130 135 140

Asp Ser Leu Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu
145 150 155 160

Asp Trp Asp Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser
165 170 175

Arg Glu Lys Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala
180 185 190

Gln Lys Val Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu
195 200 205

Lys Ser Lys Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala
210 215 220

Leu Glu Ser Trp Leu Pro Arg Asn Ile Tyr Ile Val Ala Gly Val Phe
225 230 235 240

Ile Ala Ile Ser Leu Leu Gln Ile Phe Gly Ile Phe Leu Ala Arg Thr
245 250 255

Leu Ile Ser Asp Ile Glu Ala Val Lys Ala Gly His His Phe
260 265 270

<210> 4
<211> 228
<212> PRT
<213> Homo sapiens

<400> 4
Glu Lys Gly Val Leu Ser Asp Leu Thr Lys Val Thr Arg Met His Gly
1 5 10 15

Ile Asp Pro Val Val Leu Val Leu Met Val Gly Val Val Met Phe Thr
20 25 30

Leu Gly Phe Ala Gly Cys Val Gly Ala Leu Arg Glu Asn Ile Cys Leu
35 40 45

Leu Asn Phe Phe Cys Gly Thr Ile Val Leu Ile Phe Phe Leu Glu Leu
50 55 60

Ala Val Ala Val Leu Ala Phe Leu Phe Gln Asp Trp Val Arg Asp Arg
65 70 75 80

Phe Arg Glu Phe Phe Glu Ser Asn Ile Lys Ser Tyr Arg Asp Asp Ile
85 90 95

Asp Leu Gln Asn Leu Ile Asp Ser Leu Gln Lys Ala Asn Gln Cys Cys
100 105 110

Gly Ala Tyr Gly Pro Glu Asp Trp Asp Leu Asn Val Tyr Phe Asn Cys
115 120 125

Ser Gly Ala Ser Tyr Ser Arg Glu Lys Cys Gly Val Pro Phe Ser Cys
130 135 140

Cys Val Pro Asp Pro Ala Gln Lys Val Val Asn Thr Gln Cys Gly Tyr
145 150 155 160

Asp Val Arg Ile Gln Leu Lys Ser Lys Trp Asp Glu Ser Ile Phe Thr
165 170 175

Lys Gly Cys Ile Gln Ala Leu Glu Ser Trp Leu Pro Arg Asn Ile Tyr
180 185 190

Ile Val Ala Gly Val Phe Ile Ala Ile Ser Leu Leu Gln Ile Phe Gly
195 200 205

Ile Phe Leu Ala Arg Thr Leu Ile Ser Asp Ile Glu Ala Val Lys Ala
210 215 220

Gly His His Phe
225

<210> 5
<211> 42
<212> PRT
<213> Homo sapiens

<400> 5
Met His Tyr Tyr Arg Tyr Ser Asn Ala Lys Val Ser Cys Trp Tyr Lys
1 5 10 15

Tyr Leu Leu Phe Ser Tyr Asn Ile Ile Phe Trp Leu Ala Gly Val Val
20 25 30

Phe Leu Gly Val Gly Leu Trp Ala Trp Ser
35 40

<210> 6
<211> 193
<212> PRT
<213> Homo sapiens

<400> 6
Val Gly Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys Val Gly Ala
1 5 10 15

Leu Arg Glu Asn Ile Cys Leu Leu Asn Phe Phe Cys Gly Thr Ile Val
20 25 30

Leu Ile Phe Phe Leu Glu Leu Ala Val Ala Val Leu Ala Phe Leu Phe
35 40 45

Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu Ser Asn Ile
50 55 60

Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile Asp Ser Leu
65 70 75 80

Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu Asp Trp Asp
85 90 95

Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser Arg Glu Lys
100 105 110

Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala Gln Lys Val
115 120 125

Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu Lys Ser Lys
130 135 140

Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala Leu Glu Ser
145 150 155 160

Trp Leu Pro Arg Asn Ile Tyr Ile Val Ala Gly Val Phe Ile Ala Ile
165 170 175

Ser Leu Leu Gln Ile Phe Gly Ile Phe Leu Ala Arg Thr Leu Ile Ser
180 185 190

Asp

<210> 7
<211> 23
<212> PRT
<213> Homo sapiens

<400> 7
Gly Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys Val Gly Ala Leu
1 5 10 15

Arg Glu Asn Ile Cys Leu Leu
20

<210> 8
<211> 253
<212> PRT
<213> Homo sapiens

<400> 8
Leu Leu Phe Ser Tyr Asn Ile Ile Phe Trp Leu Ala Gly Val Val Phe
1 5 10 15

Leu Gly Val Gly Leu Trp Ala Trp Ser Glu Lys Gly Val Leu Ser Asp
20 25 30

Leu Thr Lys Val Thr Arg Met His Gly Ile Asp Pro Val Val Leu Val
35 40 45

```
Leu Met Val Gly Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys Val
     50                  55                  60

Gly Ala Leu Arg Glu Asn Ile Cys Leu Leu Asn Phe Phe Cys Gly Thr
 65                  70                  75                  80

Ile Val Leu Ile Phe Phe Leu Glu Leu Ala Val Ala Val Leu Ala Phe
                 85                  90                  95

Leu Phe Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu Ser
            100                 105                 110

Asn Ile Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile Asp
        115                 120                 125

Ser Leu Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu Asp
    130                 135                 140

Trp Asp Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser Arg
145                 150                 155                 160

Glu Lys Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala Gln
                165                 170                 175

Lys Val Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu Lys
            180                 185                 190

Ser Lys Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala Leu
        195                 200                 205

Glu Ser Trp Leu Pro Arg Asn Ile Tyr Ile Val Ala Gly Val Phe Ile
    210                 215                 220

Ala Ile Ser Leu Leu Gln Ile Phe Gly Ile Phe Leu Ala Arg Thr Leu
225                 230                 235                 240

Ile Ser Asp Ile Glu Ala Val Lys Ala Gly His His Phe
                245                 250


<210> 9
<211> 9
<212> PRT
<213> Homo sapiens

<400> 9
Gly Val Pro Phe Ser Cys Cys Val Pro
  1               5


<210> 10
<211> 11
<212> PRT
<213> Homo sapiens

<400> 10
Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys
  1               5                  10
```

<210> 11
<211> 16
<212> PRT
<213> Homo sapiens

<400> 11
Met His Tyr Tyr Arg Tyr Ser Asn Ala Lys Val Ser Cys Trp Tyr Lys
1 5 10 15

<210> 12
<211> 8
<212> PRT
<213> Homo sapiens

<400> 12
Arg Glu Asn Ile Cys Leu Leu Asn
1 5

<210> 13
<211> 16
<212> PRT
<213> Homo sapiens

<400> 13
Arg Thr Leu Ile Ser Asp Ile Glu Ala Val Lys Ala Gly His His Phe
1 5 10 15

<210> 14
<211> 25
<212> PRT
<213> Homo sapiens

<400> 14
Tyr Leu Leu Phe Ser Tyr Asn Ile Ile Phe Trp Leu Ala Gly Val Val
1 5 10 15

Phe Leu Gly Val Gly Leu Trp Ala Trp
20 25

<210> 15
<211> 22
<212> PRT
<213> Homo sapiens

<400> 15
Val Leu Val Leu Met Val Gly Val Val Met Phe Thr Leu Gly Phe Ala
1 5 10 15

Gly Cys Val Gly Ala Leu
20

<210> 16

<211> 30
<212> PRT
<213> Homo sapiens

<400> 16
```
Met His Tyr Tyr Arg Tyr Ser Phe Phe Cys Gly Thr Ile Val Leu Ile
 1               5                  10                  15

Phe Phe Leu Glu Leu Ala Val Ala Val Leu Ala Phe Leu Phe
            20                  25                  30
```

<210> 17
<211> 22
<212> PRT
<213> Homo sapiens

<400> 17
```
Ile Tyr Ile Val Ala Gly Val Phe Ile Ala Ile Ser Leu Leu Gln Ile
 1               5                  10                  15

Phe Gly Ile Phe Leu Ala
            20
```

<210> 18
<211> 19
<212> PRT
<213> Homo sapiens

<400> 18
```
Ser Glu Lys Gly Val Leu Ser Asp Leu Thr Lys Val Thr Arg Met His
 1               5                  10                  15

Gly Ile Asp
```

<210> 19
<211> 117
<212> PRT
<213> Homo sapiens

<400> 19
```
Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu Ser Asn Ile
 1               5                  10                  15

Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile Asp Ser Leu
            20                  25                  30

Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu Asp Trp Asp
        35                  40                  45

Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser Arg Glu Lys
    50                  55                  60

Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala Gln Lys Val
65                  70                  75                  80
```

Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu Lys Ser Lys
85 90 95

Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala Leu Glu Ser
100 105 110

Trp Leu Pro Arg Asn
115

<210> 20
<211> 19
<212> PRT
<213> Homo sapiens

<400> 20
Glu Lys Gly Val Leu Ser Asp Leu Thr Lys Val Thr Arg Met His Gly
1 5 10 15

Ile Asp Pro

<210> 21
<211> 117
<212> PRT
<213> Homo sapiens

<400> 21
Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu Ser Asn Ile
1 5 10 15

Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile Asp Ser Leu
20 25 30

Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu Asp Trp Asp
35 40 45

Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser Arg Glu Lys
50 55 60

Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala Gln Lys Val
65 70 75 80

Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu Lys Ser Lys
85 90 95

Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala Leu Glu Ser
100 105 110

Trp Leu Pro Arg Asn
115

<210> 22
<211> 8
<212> PRT
<213> Homo sapiens

<400> 22
Arg Glu Asn Ile Cys Leu Leu Asn
1 5

<210> 23
<211> 16
<212> PRT
<213> Homo sapiens

<400> 23
Arg Thr Leu Ile Ser Asp Ile Glu Ala Val Lys Ala Gly His His Phe
1 5 10 15

<210> 24
<211> 228
<212> PRT
<213> Homo sapiens

<400> 24
Met Pro Val Lys Gly Gly Thr Lys Cys Ile Lys Tyr Leu Leu Phe Gly
1 5 10 15

Phe Asn Phe Ile Phe Trp Leu Ala Gly Ile Ala Val Leu Ala Ile Gly
20 25 30

Leu Trp Leu Arg Phe Asp Ser Gln Thr Lys Ser Ile Phe Glu Gln Glu
35 40 45

Thr Asn Asn Asn Asn Ser Ser Phe Tyr Thr Gly Val Tyr Ile Leu Ile
50 55 60

Gly Ala Gly Ala Leu Met Met Leu Val Gly Phe Leu Gly Cys Cys Gly
65 70 75 80

Ala Val Gln Glu Ser Gln Cys Met Leu Gly Leu Phe Phe Gly Phe Leu
85 90 95

Leu Val Ile Phe Ala Ile Glu Ile Ala Ala Ala Ile Trp Gly Tyr Ser
100 105 110

His Lys Asp Glu Val Ile Lys Glu Val Gln Glu Phe Tyr Lys Asp Thr
115 120 125

Tyr Asn Lys Leu Lys Thr Lys Asp Glu Pro Gln Arg Glu Thr Leu Lys
130 135 140

Ala Ile His Tyr Ala Leu Asn Cys Cys Gly Leu Ala Gly Gly Val Glu
145 150 155 160

Gln Phe Ile Ser Asp Ile Cys Pro Lys Lys Asp Val Leu Glu Thr Phe
165 170 175

Thr Val Lys Ser Cys Pro Asp Ala Ile Lys Glu Val Phe Asp Asn Lys
180 185 190

Phe His Ile Ile Gly Ala Val Gly Ile Gly Ile Ala Val Val Met Ile

195 200 205

Phe Gly Met Ile Phe Ser Met Ile Leu Cys Cys Ala Ile Arg Arg Asn
210 215 220

Arg Glu Met Val
225

<210> 25
<211> 687
<212> DNA
<213> Homo sapiens

<400> 25
atgccggtca aaggaggcac caagtgcatc aaatacctgc tgttcggatt taacttcatc 60
ttctggcttg ccgggattgc tgtccttgcc attggactat ggctccgatt cgactctcag 120
accaagagca tcttcgagca agaaactaat aataataatt ccagcttcta cacaggagtc 180
tatattctga tcggagccgg cgccctcatg atgctggtgg gcttcctggg ctgctgcggg 240
gctgtgcagg agtcccagtg catgctggga ctgttcttcg gcttcctctt ggtgatattc 300
gccattgaaa tagctgcggc catctgggga tattcccaca aggatgaggt gattaaggaa 360
gtccaggagt tttacaagga cacctacaac aagctgaaaa ccaaggatga gccccagcgg 420
gaaacgctga aagccatcca ctatgcgttg aactgctgtg gtttggctgg gggcgtggaa 480
cagtttatct cagacatctg ccccaagaag gacgtactcg aaaccttcac cgtgaagtcc 540
tgtcctgatg ccatcaaaga ggtcttcgac aataaattcc acatcatcgg cgcagtgggc 600
atcggcattg ccgtggtcat gatatttggc atgatcttca gtatgatctt gtgctgtgct 660
atccgcagga accgcgagat ggtctag 687

<210> 26
<211> 1192
<212> DNA
<213> Homo sapiens

<400> 26
cgcgcccccc agtcccgcac ccgttcggcc caggctaagt tagccctcac catgccggtc 60
aaaggaggca ccaagtgcat caaatacctg ctgttcggat ttaacttcat cttctggctt 120
gccgggattg ctgtccttgc cattggacta tggctccgat tcgactctca gaccaagagc 180
atcttcgagc aagaaactaa taataataat tccagcttct acacaggagt ctatattctg 240
atcggagccg gcgccctcat gatgctggtg ggcttcctgg gctgctgcgg ggctgtgcag 300
gagtcccagt gcatgctggg actgttcttc ggcttcctct tggtgatatt cgccattgaa 360
atagctgcgg ccatctgggg atattcccac aaggatgagg tgattaagga agtccaggag 420
ttttacaagg acacctacaa caagctgaaa accaaggatg agccccagcg ggaaacgctg 480
aaagccatcc actatgcgtt gaactgctgt ggtttggctg ggggcgtgga acagtttatc 540
tcagacatct gccccaagaa ggacgtactc gaaaccttca ccgtgaagtc ctgtcctgat 600
gccatcaaag aggtcttcga caataaattc cacatcatcg gcgcagtggg catcggcatt 660
gccgtggtca tgatatttgc catgatcttc agtatgatct tgtgctgtgc tatccgcagg 720
aaccgcgaga tggtctagag tcagcttaca tccctgagca ggaaagttta cccatgaaga 780
ttggtgggat tttttgtttg tttgttttgt tttgtttgtt gtttgttgtt tgtttttttg 840
ccactaattt tagtattcat tctgcattgc tagataaaag ctgaagttac tttatgtttg 900
tcttttaatg cttcattcaa tattgacatt tgtagttgag cggggggttt ggtttgcttg 960
gtttatattt ttcagttgtt tgtttttgct tgttatatta agcagaaatc ctgcaatgaa 1020
aggtactata tttgctagac tctagacaag atattgtaca taaaagaatt tttttgtctt 1080
taaatagata caaatgtcta tcaactttaa tcaagttgta acttatattg aagacaattt 1140
gatacataat aaaaaattat gacaatgaaa aaaaaaaaaa aaaaaaaaaa gg 1192

<210> 27
<211> 1239
<212> DNA

<213> Homo sapiens

<400> 27
ctcacaggag gagttggcgg ggagccttgg gcccctctgg cctcagccgg atttcccagc 60
caaacgcaga gagagatgcc ctggaccatc ttgctctttg cagctggctc cttggcgatc 120
ccagcaccat ccatccggct ggtgcccccg tacccaagca gccaagagga ccccatccac 180
atcgcatgca tggcccctgg gaacttcccg ggggcgaatt tcacactgta tcgagggggg 240
caggtggtcc agctcctgca ggcccccacg gaccagcgcg gggtgacatt taacctgagc 300
ggcggcagca gcaaggctcc aggggggaccc ttccactgcc agtatggagt gttaggtgag 360
ctcaaccagt cccagctgtc agacctcagc gagcccgtga acgtctcctt cccagtgccc 420
acttggatct tggtgctctc cctgagcctg gctggtgccc tcttcctcct tgctgggctg 480
gtggctgttg ccctggtggt cagaaaagtt aaactcagaa atttacagaa gaaaagagat 540
cgagaatcct gctgggccca gattaacttc gacagcacag acatgtcctt cgataactcc 600
ctgtttaccg tctccgcgaa aacgatgcca gaagaagacc cggccacctt ggatgatcac 660
tcaggcacca ctgccacccc cagcaactcc aggacccgga agaggcccac ttccacgtcc 720
tcctcgcctg agacccccga attcagcact ttccgggcct gccagtgagg ctgaggactg 780
ggggaccccct ctgtctccag gcattcgggg gcctgaggtc cctccagcta cttctggggg 840
ggctctgtca gccactttct cagggaattg gacagaggaa aggaagggga accctggcct 900
tgggattttc atcacagagg agtgggagag gggacacagg catgggcctg gcactataca 960
gacaacagga agttcccctc tcgaccttcg gctcctcagg accaccagag aaggagatgt 1020
caggacccct tcttgtcccc cagctgggcc ataagacgtc ccaggtctct gcacacccgt 1080
ggaattcctc ccttccccag tgggtttttg agcatagggt gcccttgggt gtgttgtgtg 1140
tctgcctgct ggcttgctta agttattaat tataacacgg gtcaaggtgt taaaaaaaaa 1200
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa 1239

<210> 28
<211> 693
<212> DNA
<213> Homo sapiens

<400> 28
atgccctgga ccatcttgct ctttgcagct ggctccttgg cgatcccagc accatccatc 60
cggctggtgc ccccgtaccc aagcagccaa gaggacccca tccacatcgc atgcatggcc 120
cctgggaact tcccgggggc gaatttcaca ctgtatcgag gggggcaggt ggtccagctc 180
ctgcaggccc ccacggacca gcgcggggtg acatttaacc tgagcggcgg cagcagcaag 240
gctccagggg gacccttcca ctgccagtat ggagtgttag gtgagctcaa ccagtcccag 300
ctgtcagacc tcagcgagcc cgtgaacgtc tccttcccag tgcccacttg gatcttggtg 360
ctctccctga gcctggctgg tgccctcttc ctccttgctg ggctggtggc tgttgccctg 420
gtggtcagaa aagttaaact cagaaattta cagaagaaaa gagatcgaga atcctgctgg 480
gcccagatta acttcgacag cacagacatg tccttcgata actccctgtt taccgtctcc 540
gcgaaaacga tgccagaaga agacccggcc accttggatg atcactcagg caccactgcc 600
acccccagca actccaggac ccggaagagg cccacttcca cgtcctcctc gcctgagacc 660
cccgaattca gcactttccg ggcctgccag tga 693

<210> 29
<211> 230
<212> PRT
<213> Homo sapiens

<400> 29
Met Pro Trp Thr Ile Leu Leu Phe Ala Ala Gly Ser Leu Ala Ile Pro
1 5 10 15

Ala Pro Ser Ile Arg Leu Val Pro Pro Tyr Pro Ser Ser Gln Glu Asp
20 25 30

Pro Ile His Ile Ala Cys Met Ala Pro Gly Asn Phe Pro Gly Ala Asn
35 40 45

Phe Thr Leu Tyr Arg Gly Gly Gln Val Val Gln Leu Leu Gln Ala Pro
50 55 60

Thr Asp Gln Arg Gly Val Thr Phe Asn Leu Ser Gly Gly Ser Ser Lys
65 70 75 80

Ala Pro Gly Gly Pro Phe His Cys Gln Tyr Gly Val Leu Gly Glu Leu
85 90 95

Asn Gln Ser Gln Leu Ser Asp Leu Ser Glu Pro Val Asn Val Ser Phe
100 105 110

Pro Val Pro Thr Trp Ile Leu Val Leu Ser Leu Ser Leu Ala Gly Ala
115 120 125

Leu Phe Leu Leu Ala Gly Leu Val Ala Val Ala Leu Val Val Arg Lys
130 135 140

Val Lys Leu Arg Asn Leu Gln Lys Lys Arg Asp Arg Glu Ser Cys Trp
145 150 155 160

Ala Gln Ile Asn Phe Asp Ser Thr Asp Met Ser Phe Asp Asn Ser Leu
165 170 175

Phe Thr Val Ser Ala Lys Thr Met Pro Glu Glu Asp Pro Ala Thr Leu
180 185 190

Asp Asp His Ser Gly Thr Thr Ala Thr Pro Ser Asn Ser Arg Thr Arg
195 200 205

Lys Arg Pro Thr Ser Thr Ser Ser Ser Pro Glu Thr Pro Glu Phe Ser
210 215 220

Thr Phe Arg Ala Cys Gln
225 230

<210> 30
<211> 216
<212> PRT
<213> Homo sapiens

<400> 30
Ile Pro Ala Pro Ser Ile Arg Leu Val Pro Pro Tyr Pro Ser Ser Gln
1 5 10 15

Glu Asp Pro Ile His Ile Ala Cys Met Ala Pro Gly Asn Phe Pro Gly
20 25 30

Ala Asn Phe Thr Leu Tyr Arg Gly Gly Gln Val Val Gln Leu Leu Gln
35 40 45

Ala Pro Thr Asp Gln Arg Gly Val Thr Phe Asn Leu Ser Gly Gly Ser
50 55 60

Ser Lys Ala Pro Gly Gly Pro Phe His Cys Gln Tyr Gly Val Leu Gly
65 70 75 80

Glu Leu Asn Gln Ser Gln Leu Ser Asp Leu Ser Glu Pro Val Asn Val
85 90 95

Ser Phe Pro Val Pro Thr Trp Ile Leu Val Leu Ser Leu Ser Leu Ala
100 105 110

Gly Ala Leu Phe Leu Leu Ala Gly Leu Val Ala Val Ala Leu Val Val
115 120 125

Arg Lys Val Lys Leu Arg Asn Leu Gln Lys Lys Arg Asp Arg Glu Ser
130 135 140

Cys Trp Ala Gln Ile Asn Phe Asp Ser Thr Asp Met Ser Phe Asp Asn
145 150 155 160

Ser Leu Phe Thr Val Ser Ala Lys Thr Met Pro Glu Glu Asp Pro Ala
165 170 175

Thr Leu Asp Asp His Ser Gly Thr Thr Ala Thr Pro Ser Asn Ser Arg
180 185 190

Thr Arg Lys Arg Pro Thr Ser Thr Ser Ser Ser Pro Glu Thr Pro Glu
195 200 205

Phe Ser Thr Phe Arg Ala Cys Gln
210 215

<210> 31
<211> 14
<212> PRT
<213> Homo sapiens

<400> 31
Met Pro Trp Thr Ile Leu Leu Phe Ala Ala Gly Ser Leu Ala
1 5 10

<210> 32
<211> 102
<212> PRT
<213> Homo sapiens

<400> 32
Ile Pro Ala Pro Ser Ile Arg Leu Val Pro Pro Tyr Pro Ser Ser Gln
1 5 10 15

Glu Asp Pro Ile His Ile Ala Cys Met Ala Pro Gly Asn Phe Pro Gly
20 25 30

Ala Asn Phe Thr Leu Tyr Arg Gly Gly Gln Val Val Gln Leu Leu Gln
35 40 45

Ala Pro Thr Asp Gln Arg Gly Val Thr Phe Asn Leu Ser Gly Gly Ser
50 55 60

Ser Lys Ala Pro Gly Gly Pro Phe His Cys Gln Tyr Gly Val Leu Gly
65 70 75 80

```
Glu Leu Asn Gln Ser Gln Leu Ser Asp Leu Ser Glu Pro Val Asn Val
                85                  90                  95

Ser Phe Pro Val Pro Thr
            100


<210> 33
<211> 25
<212> PRT
<213> Homo sapiens

<400> 33
Trp Ile Leu Val Leu Ser Leu Ser Leu Ala Gly Ala Leu Phe Leu Leu
  1               5                  10                  15

Ala Gly Leu Val Ala Val Ala Leu Val
             20                  25


<210> 34
<211> 89
<212> PRT
<213> Homo sapiens

<400> 34
Val Arg Lys Val Lys Leu Arg Asn Leu Gln Lys Lys Arg Asp Arg Glu
  1               5                  10                  15

Ser Cys Trp Ala Gln Ile Asn Phe Asp Ser Thr Asp Met Ser Phe Asp
             20                  25                  30

Asn Ser Leu Phe Thr Val Ser Ala Lys Thr Met Pro Glu Glu Asp Pro
         35                  40                  45

Ala Thr Leu Asp Asp His Ser Gly Thr Thr Ala Thr Pro Ser Asn Ser
     50                  55                  60

Arg Thr Arg Lys Arg Pro Thr Ser Thr Ser Ser Ser Pro Glu Thr Pro
 65                  70                  75                  80

Glu Phe Ser Thr Phe Arg Ala Cys Gln
                85


<210> 35
<211> 7
<212> PRT
<213> Homo sapiens

<400> 35
Gln Ala Gly Ser Val Tyr Val
  1               5


<210> 36
<211> 1608
<212> DNA
```

<213> Homo sapiens

<400> 36
gcactctaac gaaatcaaaa gtagaaatta tgtgatacac tacagatggg tttatgggag 60
agagacagca caaacccccа aacttgggcc ttcagacaaa gaaaggcttt ggccctacat 120
aagagtgttg ccaactaaat atgtatttgc tctaaactta ataggcttag ggaggcatcg 180
tgtatgtata aactatacat acatacatac atatgtgttt atacatacac aatgcctata 240
atgattcttc acttaatttt tcaaatttct catcaagtat tggtcttaat tgttcctttt 300
aagagtgctt ctgtaagtat taaatctaac ttatatattc cattaatttg taatttaatt 360
gcgtgtccaa tgtacagcag taacaatcag aatcttcaca aaggccagtg ccattttgta 420
aaatcttttt aaaggattaa ttcagttatg ttataattaa gtataaacat cgatatgaat 480
actttcaagt ctccagttca tattactata tgtcaagtgg agaattaatt ttttttaat 540
ttggtaaact agtagtttag atagcttcag tgtaggtagt ccctaaaaat atctgggtag 600
actgtgaaat tgtctcatat gccttgataa tttcagtttt taaaaggagt agcttaatat 660
ttctatttc tgttctttta ctcagacatt attattatta ttattattat tattattatt 720
ttttattatt attattttac aaattgtatt tagtatgtgt ttttgaaaca caaaactggc 780
atgctggtgt atctggtgaa tttactatta agcccaggtc tgcaccagat gggcaggtga 840
aaatgtgatt atgaggtgtg tggcgttata tccatctctg taatacttga ctctccaggt 900
gacttgtacg tctgttacat tcatgaattc actttacggt tattcttaca gatactgaaa 960
gaagcaacct tacaccacgc aaggtctatg gaaaacaaaa acttccctcc tttaacacaa 1020
tcacaataaa tgtttcataa caaccagata aatagaaaca tgtcagcatt ttccccggca 1080
ggttcagaag tttgatgctg aaatactggt ttgcacatct agaccaagac taaagtgagt 1140
ttgctatttc aaaggaaaat gcaaattgag gcttaggctg cacctttctg tatcactgtt 1200
ttggtaagtc cacatggggc agacaaagcc aaacaccagc taggtggtag aattccccaa 1260
atattcatct gtgctgaaac cttcaaaagt caggtgcctg gagagctcat ttaatgaaag 1320
ggttatctct gccaaccagt cagtagttga tatttttttc accagaagtt gtgaaatatt 1380
atttatgtat gtagaagaca ggggcttcct gccgttgttg tttacttgtg tgtttgggca 1440
agtttcctcc tttccagaaa gcaagtttc cataaaagca actttgcaat ggccctactc 1500
cttttgtttg ataccccagg caaaaagctt ttcatggtat tgaaataatt ttcatttaaa 1560
aacagggagc tcagatgaac cagtccttta ttaaaaaaaa aaaaaaa 1608

<210> 37
<211> 249
<212> DNA
<213> Homo sapiens

<400> 37
atgtataaac tatacataca tacatacata tgtgtttata catacacaat gcctataatg 60
attcttcact taatttttca aatttctcat caagtattgg tcttaattgt tccttttaag 120
agtgcttctg taagtattaa atctaactta tatattccat taatttgtaa tttaattgcg 180
tgtccaatgt acagcagtaa caatcagaat cttcacaaag gccagtgcca ttttgtaaaa 240
tctttttaa 249

<210> 38
<211> 82
<212> PRT
<213> Homo sapiens

<400> 38
Met Tyr Lys Leu Tyr Ile His Thr Tyr Ile Cys Val Tyr Thr Tyr Thr
1 5 10 15

Met Pro Ile Met Ile Leu His Leu Ile Phe Gln Ile Ser His Gln Val
20 25 30

Leu Val Leu Ile Val Pro Phe Lys Ser Ala Ser Val Ser Ile Lys Ser
35 40 45

Asn Leu Tyr Ile Pro Leu Ile Cys Asn Leu Ile Ala Cys Pro Met Tyr
50 55 60

Ser Ser Asn Asn Gln Asn Leu His Lys Gly Gln Cys His Phe Val Lys
65 70 75 80

Ser Phe

<210> 39
<211> 40
<212> PRT
<213> Homo sapiens

<400> 39
Ser Val Ser Ile Lys Ser Asn Leu Tyr Ile Pro Leu Ile Cys Asn Leu
1 5 10 15

Ile Ala Cys Pro Met Tyr Ser Ser Asn Asn Gln Asn Leu His Lys Gly
20 25 30

Gln Cys His Phe Val Lys Ser Phe
35 40

<210> 40
<211> 42
<212> PRT
<213> Homo sapiens

<400> 40
Met Tyr Lys Leu Tyr Ile His Thr Tyr Ile Cys Val Tyr Thr Tyr Thr
1 5 10 15

Met Pro Ile Met Ile Leu His Leu Ile Phe Gln Ile Ser His Gln Val
20 25 30

Leu Val Leu Ile Val Pro Phe Lys Ser Ala
35 40

<210> 41
<211> 7
<212> PRT
<213> Homo sapiens

<400> 41
Ser Val Ser Ile Lys Ser Asn
1 5

<210> 42
<211> 19
<212> PRT
<213> Homo sapiens

<400> 42

Leu Tyr Ile Pro Leu Ile Cys Asn Leu Ile Ala Cys Pro Met Tyr Ser
1 5 10 15

Ser Asn Asn

<210> 43
<211> 16
<212> PRT
<213> Homo sapiens

<400> 43
Asn Asn Gln Asn Leu His Lys Gly Gln Cys His Phe Val Lys Ser Phe
1 5 10 15

<210> 44
<211> 1338
<212> DNA
<213> Homo sapiens

<400> 44
gtggatcttc acagtgcacg gctttgggcg gcccctgctg ctgtcggccc tgcacatgct 60
ggtggcagcc ctggcatgcc accgggggc acggcgcccc atgccaggcg gcactcgctg 120
ccgagtccta ctgctcagtc tcacctttgg cacgtccatg gcctgcggca acgtgggcct 180
aagggctgtg cccctggacc tggcacaact ggttactacc accacacctc tgttcaccct 240
ggccctgtcg gcgctgctgc tgggccgccg ccaccacccg cttcagttgg ccgccatggg 300
tccgctctgc ctgggggccg cctgcagcct ggctggagag ttccggacac cccctaccgg 360
ctgtggcttc ctgctcgcag ccacctgcct ccgcggactc aagtcggttc agcaaaacag 420
ggtctggctc tgtcacccag gctgcattgg tgagatctca gctcaataca gcctccgcat 480
cctgggttca agtgattctt ctgcctcagc ctcccaagtg ccctgctgca ggaggagagg 540
ctggacgcgg tgaccctgct ttacgccacc tcgctgccca gcttctgcct gctggcgggt 600
gcagccctgg tgctggaggc tggcgttgcc ccaccgccca ctgctggcga ctctcgcctc 660
tgggcctgca tcctgctcag ctgcctcctg tctgttctct ataacctggc cagcttctcc 720
ctgctggccc tcacctctgc cctcaccgtc cacgtcctgg gcaacctcac cgtggtgggc 780
aacctcatcc tgtcccggct gttgtttggc agccgcctca gtgccctcag ctacgtgggc 840
atcgcactca ctctttcagg aatgttcctt taccacaact gcgagttcgt ggcctcctgg 900
gctgcccgtc gggggctgtg gcggagggac cagcccagca agggtctttg agacctgggg 960
gatctcagga gccacctggg atggccctgg cctgaatcca gcctccgctg tggccataga 1020
aggaatggag aacagggctg ggcatggtgg ctcacgccta taatcccagc acttccagag 1080
tccgaggtgg gtggatcacc tgaggccagg agttcgagac cagcctggct aacatggcaa 1140
aacctcatct ctactaaaaa tagaaaaatt agctgggcat ggtggcgcgt gcctatagtc 1200
ccagctacat gggaggctaa ggtgggagga tcacttgagc cctggagatc gaggctgcag 1260
taagccaaga tcgcatgcta ctgcactcca gcctgggaga cagagcgaga cgctgtctca 1320
attaaaaaaa aaaaaaaa 1338

<210> 45
<211> 498
<212> DNA
<213> Homo sapiens

<400> 45
atgctggtgg cagccctggc atgccaccgg ggggcacggc gccccatgcc aggcggcact 60
cgctgccgag tcctactgct cagtctcacc tttggcacgt ccatggcctg cggcaacgtg 120
ggcctaaggg ctgtgcccct ggacctggca caactggtta ctaccaccac acctctgttc 180
accctggccc tgtcggcgct gctgctgggc cgccgccacc acccgcttca gttggccgcc 240
atgggtccgc tctgcctggg ggccgcctgc agcctggctg gagagttccg gacacccccct 300

accggctgtg gcttcctgct cgcagccacc tgcctccgcg gactcaagtc ggttcagcaa 360
aacagggtct ggctctgtca cccaggctgc attggtgaga tctcagctca atacagcctc 420
cgcatcctgg gttcaagtga ttcttctgcc tcagcctccc aagtgccctg ctgcaggagg 480
agaggctgga cgcggtga 498

<210> 46
<211> 165
<212> PRT
<213> Homo sapiens

<400> 46
Met Leu Val Ala Ala Leu Ala Cys His Arg Gly Ala Arg Arg Pro Met
1 5 10 15

Pro Gly Gly Thr Arg Cys Arg Val Leu Leu Leu Ser Leu Thr Phe Gly
20 25 30

Thr Ser Met Ala Cys Gly Asn Val Gly Leu Arg Ala Val Pro Leu Asp
35 40 45

Leu Ala Gln Leu Val Thr Thr Thr Thr Pro Leu Phe Thr Leu Ala Leu
50 55 60

Ser Ala Leu Leu Leu Gly Arg Arg His His Pro Leu Gln Leu Ala Ala
65 70 75 80

Met Gly Pro Leu Cys Leu Gly Ala Ala Cys Ser Leu Ala Gly Glu Phe
85 90 95

Arg Thr Pro Pro Thr Gly Cys Gly Phe Leu Leu Ala Ala Thr Cys Leu
100 105 110

Arg Gly Leu Lys Ser Val Gln Gln Asn Arg Val Trp Leu Cys His Pro
115 120 125

Gly Cys Ile Gly Glu Ile Ser Ala Gln Tyr Ser Leu Arg Ile Leu Gly
130 135 140

Ser Ser Asp Ser Ser Ala Ser Ala Ser Gln Val Pro Cys Cys Arg Arg
145 150 155 160

Arg Gly Trp Thr Arg
165

<210> 47
<211> 36
<212> PRT
<213> Homo sapiens

<400> 47
Met Leu Val Ala Ala Leu Ala Cys His Arg Gly Ala Arg Arg Pro Met
1 5 10 15

Pro Gly Gly Thr Arg Cys Arg Val Leu Leu Leu Ser Leu Thr Phe Gly
20 25 30

Thr Ser Met Ala

35

<210> 48
<211> 129
<212> PRT
<213> Homo sapiens

<400> 48
Cys Gly Asn Val Gly Leu Arg Ala Val Pro Leu Asp Leu Ala Gln Leu
1 5 10 15

Val Thr Thr Thr Thr Pro Leu Phe Thr Leu Ala Leu Ser Ala Leu Leu
20 25 30

Leu Gly Arg Arg His His Pro Leu Gln Leu Ala Ala Met Gly Pro Leu
35 40 45

Cys Leu Gly Ala Ala Cys Ser Leu Ala Gly Glu Phe Arg Thr Pro Pro
50 55 60

Thr Gly Cys Gly Phe Leu Leu Ala Ala Thr Cys Leu Arg Gly Leu Lys
65 70 75 80

Ser Val Gln Gln Asn Arg Val Trp Leu Cys His Pro Gly Cys Ile Gly
85 90 95

Glu Ile Ser Ala Gln Tyr Ser Leu Arg Ile Leu Gly Ser Ser Asp Ser
100 105 110

Ser Ala Ser Ala Ser Gln Val Pro Cys Cys Arg Arg Arg Gly Trp Thr
115 120 125

Arg

<210> 49
<211> 34
<212> PRT
<213> Homo sapiens

<400> 49
Cys Gly Asn Val Gly Leu Arg Ala Val Pro Leu Asp Leu Ala Gln Leu
1 5 10 15

Val Thr Thr Thr Thr Pro Leu Phe Thr Leu Ala Leu Ser Ala Leu Leu
20 25 30

Leu Gly

<210> 50
<211> 17
<212> PRT
<213> Homo sapiens

<400> 50
Leu Ala Ala Met Gly Pro Leu Cys Leu Gly Ala Ala Cys Ser Leu Ala
1   5   10   15

Gly

<210> 51
<211> 71
<212> PRT
<213> Homo sapiens

<400> 51
Glu Phe Arg Thr Pro Pro Thr Gly Cys Gly Phe Leu Leu Ala Ala Thr
1   5   10   15

Cys Leu Arg Gly Leu Lys Ser Val Gln Gln Asn Arg Val Trp Leu Cys
            20   25   30

His Pro Gly Cys Ile Gly Glu Ile Ser Ala Gln Tyr Ser Leu Arg Ile
        35   40   45

Leu Gly Ser Ser Asp Ser Ser Ala Ser Ala Ser Gln Val Pro Cys Cys
    50   55   60

Arg Arg Arg Gly Trp Thr Arg
65   70

<210> 52
<211> 983
<212> DNA
<213> Homo sapiens

<400> 52
cggacgcgtg ggtgtttgat ctgccttttg tgcgtggggt gggagttagg taggaatctt 60
aaagtggaga gccagtttct tcccaaatta ctgacctaac ccatccttaa cccccagttc 120
aaggccacct ttgtgatagt gaagcttcca catgctcact cagccccttc tgctctctct 180
tcttctctac tgtgcatgtc ggcttgtact tttgccagtt tctctaaaga cacaaccaga 240
ggtggggtgg ctgtgtgtgc acaacttcaa ctttacatgt ggggctgagt ccctatgttg 300
tatatccttg tgcaaaagca caatatgtta attgctatag cttttaaaaa aataattaat 360
agtttttcat aatcaaattt tcttgctttt ttgtttttttc aaaaaagcat acttttattg 420
aagaataaac cccttatata tgtacactta tttataacta tgaaccatga actaggatag 480
aaatgcattg tgtatattac aaaacataac aaaaataata ggggtaggga ggtgcagatg 540
ttggtcaaag gatataaacc tgcagttcta tgatgaataa gttctggaca tctggaatac 600
agcatggtga ctatacttag taatactata ttgtacactt gaagcttact gaaagagtaa 660
atctcaagtg ttctcaccac acaaacccaa aggtaactat gttctcacca cacaaaccca 720
aagggaacta tgtattaatt agcttgattg tggtaaccat ttcacaatgt atacatttgc 780
caaaacatta tgttgtatac ctggaatata taattttatt tatcaattat acctcaataa 840
agctgaaaga ggggattact aattcccaca aaatacagat ttaacaaaaa cttttattca 900
acaaacagtg ctatgaagtt gtaaattgga aacaaaagaa ataaaatttc atccacagtc 960
ttctcatcaa aaaaaaaaaa aaa 983

<210> 53
<211> 180
<212> DNA
<213> Homo sapiens

<400> 53
atgctcactc agccccttct gctctctctt cttctctact gtgcatgtcg gcttgtactt 60
ttgccagttt ctctaaagac acaaccagag gtggggtggc tgtgtgtgca caacttcaac 120
tttacatgtg ggctgagtc cctatgttgt atatccttgt gcaaaagcac aatatgttaa 180

<210> 54
<211> 59
<212> PRT
<213> Homo sapiens

<400> 54
Met Leu Thr Gln Pro Leu Leu Leu Ser Leu Leu Leu Tyr Cys Ala Cys
1 5 10 15

Arg Leu Val Leu Leu Pro Val Ser Leu Lys Thr Gln Pro Glu Val Gly
20 25 30

Trp Leu Cys Val His Asn Phe Asn Phe Thr Cys Gly Ala Glu Ser Leu
35 40 45

Cys Cys Ile Ser Leu Cys Lys Ser Thr Ile Cys
50 55

<210> 55
<211> 32
<212> PRT
<213> Homo sapiens

<400> 55
Gln Pro Glu Val Gly Trp Leu Cys Val His Asn Phe Asn Phe Thr Cys
1 5 10 15

Gly Ala Glu Ser Leu Cys Cys Ile Ser Leu Cys Lys Ser Thr Ile Cys
20 25 30

<210> 56
<211> 27
<212> PRT
<213> Homo sapiens

<400> 56
Met Leu Thr Gln Pro Leu Leu Leu Ser Leu Leu Leu Tyr Cys Ala Cys
1 5 10 15

Arg Leu Val Leu Leu Pro Val Ser Leu Lys Thr
20 25

<210> 57
<211> 973
<212> DNA
<213> Homo sapiens

<400> 57
ctgcacatgt atttgtttga tctgcctttt gtgcgtgggg tgggagttag gtaggaatct 60
taaagtggag agccagtttc ttcccaaatt actgacctaa cccatcctta accccagtt 120
caaggccacc tttgtgatag tgaagcttcc acatgctcac tcagcccctt ctgctctctc 180
ttcttctcta ctgtgcatgt cggcttgtac ttttgccagt ttctctaaag acacaaccag 240
aggtggggtg gctgtgtgtg cacaacttca actttacatg tggggctgag tccctatgtt 300
gtatatcctt gtgcaaaagc acaatatgtt aattgctata gcttttaaaa aaataattaa 360
tagtttttca taatcaaatt ttcttgcttt tttgtttttt caaaaaagca tacttttatt 420
gaagaataaa ccccttatat atgtacactt atttataact atgaaccatg aactaggata 480
gaaatgcatt gtgtatatta caaaacataa caaaaataat aggggtaggg aggtgcagat 540
gttggtcaaa ggatataaac ctgcagttct atgatgaata agttctggac atctggaata 600
cagcatggtg actatactta gtaatactat attgtacact tgaagcttac tgaaagagta 660
aatctcaagt gttctcacca cacaaaccca aaggtaacta tgttctcacc acacaaaccc 720
aaagggaact atgtattaat tagcttgatt gtggtaacca tttcacaatg tatacatttg 780
ccaaaacatt atgttgtata cctggaatat ataattttat ttatcaatta tacctcaata 840
aagctgaaag aggggattac taattcccac aaaatacaga tttaacaaaa acttttattc 900
aacaaacagt gctatgaagt tgtaaattgg aaacaaaaga aataaaattt catccacagt 960
cttctcatca aaa 973

<210> 58
<211> 1119
<212> DNA
<213> Homo sapiens

<400> 58
ggaaccctcc tgtcccccctg ccctctctgc tcacatcatc tgctgcccta gactttggag 60
gcttgaattt ctgcacgagg tgactgccaa caaatgacaa gagcagccat ttagtgagca 120
cgtaattcat ttaatggtgc tgaacacaga gcaagtgctg catgagtatt ttggttaggg 180
ttcacttgta ccttttaggt ttagctctta tgcaaagcct gtggttcaga tccatgtgtc 240
accctcaggt cactacgtcc cactgcagca ggtatgggga gaatcataac cataacacct 300
tcccttgcag tgaatttctc tctcatattt gtctttagtt tgaaccacat aataaatcta 360
taagcgtatt atagttccta ttctatagat gaggagactg aggcacacta agggaaaaag 420
tgacaaggaa gagactagag gctacatctg atttacacc aagtattcat ccccacacaat 480
gaatagcaac caccggatgt ttttcaagat ttgagtgaag gccagaatca aactgaagac 540
tatgtgtgtg tgtgtatgtg tgttcatatt caaaaacact acatacacgc tacatgtatt 600
atatatatta tataatatac acacataata acatataaca tataagtata taataggtaa 660
ttatatattt ttatattaat aatataatat atattatata tataatatac acacatatac 720
atactttcta taaatctact ccaaaagctt caaggtctcc caaatatcac atgacccata 780
gctaaggcaa tcctgaaacc tggctgcttg ggctagcctg gggacacacg aggagtcagg 840
aactgggctg tccttccttg caatgctctc tccttctgcc aaaacttatc tttctgatct 900
gcctttcccc ccaggatcca gggtggtcct aggaaaccca aggaaacgct tccagctgga 960
gtgctcggag gtgtaggaca ttgttctctt cccttcccgg gtcctgttgt tttagaacct 1020
aatcaataaa aattaagctg gtaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1080
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa 1119

<210> 59
<211> 177
<212> DNA
<213> Homo sapiens

<400> 59
atgagtattt tggttagggt tcacttgtac cttttaggtt tagctcttat gcaaagcctg 60
tggttcagat ccatgtgtca ccctcaggtc actacgtccc actgcagcag gtatggggag 120
aatcataacc ataacacctt cccttgcagt gaatttctct ctcatatttg tctttag 177

<210> 60
<211> 58
<212> PRT

<213> Homo sapiens

<400> 60
Met Ser Ile Leu Val Arg Val His Leu Tyr Leu Leu Gly Leu Ala Leu
1 5 10 15

Met Gln Ser Leu Trp Phe Arg Ser Met Cys His Pro Gln Val Thr Thr
20 25 30

Ser His Cys Ser Arg Tyr Gly Glu Asn His Asn His Asn Thr Phe Pro
35 40 45

Cys Ser Glu Phe Leu Ser His Ile Cys Leu
50 55

<210> 61
<211> 32
<212> PRT
<213> Homo sapiens

<400> 61
His Pro Gln Val Thr Thr Ser His Cys Ser Arg Tyr Gly Glu Asn His
1 5 10 15

Asn His Asn Thr Phe Pro Cys Ser Glu Phe Leu Ser His Ile Cys Leu
20 25 30

<210> 62
<211> 26
<212> PRT
<213> Homo sapiens

<400> 62
Met Ser Ile Leu Val Arg Val His Leu Tyr Leu Leu Gly Leu Ala Leu
1 5 10 15

Met Gln Ser Leu Trp Phe Arg Ser Met Cys
20 25

<210> 63
<211> 1386
<212> DNA
<213> Homo sapiens

<400> 63
cggacgcgtg gggggagaa tagtcacagc ctcctcaaag ggttgtgaaa atcaaatgtg 60
ataatttgtg gaaagccctt agcagtggcc tggcacaaaa caaatgctca gtggatggaa 120
gctgcctatt attattgtcg ttgttgttgt ttgccatgac tgctctgggc cgggggtaga 180
gctagcatcc gggcatgtac gagggaagag ggaggcaggc ctctattcaa aggcagaaat 240
tcctttaaga ttgtggtctg ctgggtttca gggagtgtct gtgttgtttg tttttgtttg 300
tttgtttgtt ttgagacagg gtctcgctct gtcacccagg ctggagtgta gtggtgcagt 360
cttggctcac tgcaacctcc acctcctggg ctcaagcgat tctcatgcct cagcctcccg 420

agtagctggg actacaggtg tgtgccacta tgcctggcta atttttgtat tttttgtaga 480
gacggggttt tgccatgttg cccaggctgg aagtgtctat gtttaactgc atcttataaa 540
ccagcaacaa gttttctact gggaattaga atggtgcata cacaatgtat tattatcact 600
gtcagatgag catgcttgaa tgtagcatga ctgcctcttt ttgcttttcc tagaggtttt 660
tttttgctt gttactcatc tgttgaccta cctgggggaa gtagcaccct tgcatttcaa 720
aaataaaatt gatggcatta caaatggaat agaaaccatt tttaaaatat tttcagttct 780
ctttcaaaat tacctatttt atcttttata tatttgaaca tattaagcat agttatttta 840
aagttcagtt agtcccttat atggagcctc catgagtctg tttctagtgc ctgttgtttc 900
tcttgttttt tgttagtttt gtcctgtctc ctcccatact tggcaatgtt ttaattgagt 960
acaagtaatt tcaggtctag gatggagtta tcttcctcta tgaggattat gtttacatct 1020
ggcaggtggc tggggatgca gtgagccaga tctctttcat ctacttgtag gggatgagag 1080
gatttgagtc actttaaggg ctgacctact tctgtttcat tcatattcat agggtgaagc 1140
ccactggggc tccaacccaa agcatgaaac atttgccagg tagccttcct gttggtgggc 1200
tttgcagtgt cttctagaat cagcagaccc cctacagaaa acgtagtccc agatgccatg 1260
catatgtctc tgagttgcct tcttttccca gatcttggcc ttgtaattct tcactgcttt 1320
gtttcctgcc caatactctc aattagactt aaatatatat atatacgtac aaaaaaaaaa 1380
aaaaaa 1386

<210> 64
<211> 423
<212> DNA
<213> Homo sapiens

<400> 64
atgctcagtg gatggaagct gcctattatt attgtcgttg ttgttgtttg ccatgactgc 60
tctgggccgg gggtagagct agcatccggg catgtacgag ggaagaggga ggcaggcctc 120
tattcaaagg cagaaattcc tttaagattg tggtctgctg ggtttcaggg agtgtctgtg 180
ttgtttgttt ttgtttgttt gtttgttttg agacagggtc tcgctctgtc acccaggctg 240
gagtgtagtg gtgcagtctt ggctcactgc aacctccacc tcctgggctc aagcgattct 300
catgcctcag cctcccgagt agctgggact acaggtgtgt gccactatgc ctggctaatt 360
tttgtatttt ttgtagagac ggggttttgc catgttgccc aggctggaag tgtctatgtt 420
taa 423

<210> 65
<211> 140
<212> PRT
<213> Homo sapiens

<400> 65
Met Leu Ser Gly Trp Lys Leu Pro Ile Ile Ile Val Val Val Val Val
1 5 10 15

Cys His Asp Cys Ser Gly Pro Gly Val Glu Leu Ala Ser Gly His Val
20 25 30

Arg Gly Lys Arg Glu Ala Gly Leu Tyr Ser Lys Ala Glu Ile Pro Leu
35 40 45

Arg Leu Trp Ser Ala Gly Phe Gln Gly Val Ser Val Leu Phe Val Phe
50 55 60

Val Cys Leu Phe Val Leu Arg Gln Gly Leu Ala Leu Ser Pro Arg Leu
65 70 75 80

Glu Cys Ser Gly Ala Val Leu Ala His Cys Asn Leu His Leu Leu Gly
85 90 95

Ser Ser Asp Ser His Ala Ser Ala Ser Arg Val Ala Gly Thr Thr Gly

```
            100                 105                 110

Val Cys His Tyr Ala Trp Leu Ile Phe Val Phe Phe Val Glu Thr Gly
        115                 120                 125

Phe Cys His Val Ala Gln Ala Gly Ser Val Tyr Val
    130                 135                 140


<210> 66
<211> 20
<212> PRT
<213> Homo sapiens

<400> 66
Met Leu Ser Gly Trp Lys Leu Pro Ile Ile Ile Val Val Val Val Val
  1               5                  10                  15

Cys His Asp Cys
            20


<210> 67
<211> 120
<212> PRT
<213> Homo sapiens

<400> 67
Ser Gly Pro Gly Val Glu Leu Ala Ser Gly His Val Arg Gly Lys Arg
  1               5                  10                  15

Glu Ala Gly Leu Tyr Ser Lys Ala Glu Ile Pro Leu Arg Leu Trp Ser
            20                  25                  30

Ala Gly Phe Gln Gly Val Ser Val Leu Phe Val Phe Val Cys Leu Phe
        35                  40                  45

Val Leu Arg Gln Gly Leu Ala Leu Ser Pro Arg Leu Glu Cys Ser Gly
    50                  55                  60

Ala Val Leu Ala His Cys Asn Leu His Leu Leu Gly Ser Ser Asp Ser
65                  70                  75                  80

His Ala Ser Ala Ser Arg Val Ala Gly Thr Thr Gly Val Cys His Tyr
                85                  90                  95

Ala Trp Leu Ile Phe Val Phe Phe Val Glu Thr Gly Phe Cys His Val
            100                 105                 110

Ala Gln Ala Gly Ser Val Tyr Val
        115                 120


<210> 68
<211> 21
<212> PRT
<213> Homo sapiens
```

<400> 68
Leu Trp Ser Ala Gly Phe Gln Gly Val Ser Val Leu Phe Val Phe Val
1 5 10 15

Cys Leu Phe Val Leu
20

<210> 69
<211> 18
<212> PRT
<213> Homo sapiens

<400> 69
Tyr Ala Trp Leu Ile Phe Val Phe Phe Val Glu Thr Gly Phe Cys His
1 5 10 15

Val Ala

<210> 70
<211> 45
<212> PRT
<213> Homo sapiens

<400> 70
Arg Gln Gly Leu Ala Leu Ser Pro Arg Leu Glu Cys Ser Gly Ala Val
1 5 10 15

Leu Ala His Cys Asn Leu His Leu Leu Gly Ser Ser Asp Ser His Ala
20 25 30

Ser Ala Ser Arg Val Ala Gly Thr Thr Gly Val Cys His
35 40 45

<210> 71
<211> 58
<212> PRT
<213> Homo sapiens

<400> 71
Arg Leu Glu Cys Ser Gly Ala Val Leu Ala His Cys Asn Leu His Leu
1 5 10 15

Leu Gly Ser Ser Asp Ser His Ala Ser Ala Ser Arg Val Ala Gly Thr
20 25 30

Thr Gly Val Cys His Tyr Ala Trp Leu Ile Phe Val Phe Phe Val Glu
35 40 45

Thr Gly Phe Cys His Val Ala Gln Ala Gly
50 55

<210> 72
<211> 35

<212> PRT
<213> Homo sapiens

<400> 72
Ala Ser Ala Ser Arg Val Ala Gly Thr Thr Gly Val Cys His Tyr Ala
1   5   10   15

Trp Leu Ile Phe Val Phe Phe Val Glu Thr Gly Phe Cys His Val Ala
            20   25   30

Gln Ala Gly
        35

<210> 73
<211> 1778
<212> DNA
<213> Homo sapiens

<400> 73
cgactttcag tccccgacgc gccccgccca acccctacga tgaagagggc gtccgctgga 60
gggagccggc tgctggcatg ggtgctgtgg ctgcaggcct ggcaggtggc agccccatgc 120
ccaggtgcct gcgtatgcta caatgagccc aaggtgacga caagctgccc ccagcagggc 180
ctgcaggctg tgcccgtggg catccctgct gccagccagc gcatcttcct gcacggcaac 240
cgcatctcgc atgtgccagc tgccagcttc cgtgcctgcc gcaacctcac catcctgtgg 300
ctgcactcga atgtgctggc ccgaattgat gcggctgcct tcactggcct ggccctcctg 360
gagcagctgg acctcagcga taatgcacag ctccggtctg tggaccctgc cacattccac 420
ggcctgggcc gcgtacacac gctgcacctg gaccgctgcg gcctgcagga gctgggcccg 480
gggctgttcc gcggcctggc tgccctgcag tacctctacc tgcaggacaa cgcgctgcag 540
gcactgcctg atgacacctt ccgcgacctg ggcaacctca cacacctctt cctgcacggc 600
aaccgcatct ccagcgtgcc cgagcgcgcc ttccgtgggc tgcacagcct cgaccgtctc 660
ctactgcacc agaaccgcgt ggcccatgtg cacccgcatg ccttccgtga ccttggccgc 720
ctcatgacac tctatctgtt tgccaacaat ctatcagcgc tgcccactga ggccctggcc 780
cccctgcgtg ccctgcagta cctgaggctc aacgacaacc cctgggtgtg tgactgccgg 840
gcacgcccac tctgggcctg gctgcagaag ttccgcggct cctcctccga ggtgccctgc 900
agcctcccgc aacgcctggc tggccgtgac ctcaaacgcc tagctgccaa tgacctgcag 960
ggctgcgctg tggccaccgg cccttaccat cccatctgga ccggcagggc caccgatgag 1020
gagccgctgg ggcttcccaa gtgctgccag ccagatgccg ctgacaaggc ctcagtactg 1080
gagcctggaa gaccagcttc ggcaggcaat gcgctgaagg gacgcgtgcc gcccggtgac 1140
agcccgccgg gcaacggctc tggcccacgg cacatcaatg actcaccctt tgggactctg 1200
cctggctctg ctgagccccc gctcactgca gtgcggcccg agggctccga gccaccaggg 1260
ttccccacct cgggccctcg ccggaggcca ggctgttcac gcaagaaccg cacccgcagc 1320
cactgccgtc tgggccaggc aggcagcggg ggtggcggga ctggtgactc agaaggctca 1380
ggtgccctac ccagcctcac ctgcagcctc acccccctgg gcctggcgct ggtgctgtgg 1440
acagtgcttg ggccctgctg acccccagcg gacacaagag cgtgctcagc agccaggtgt 1500
gtgtacatac ggggtctctc tccacgccgc caagccagcc gggcggcgga cccgtggggc 1560
aggccaggcc aggtcctccc tgatggacgc ctgccgcccg ccacccccat ctccacccca 1620
tcatgtttac agggttcggc ggcagcgttt gttccagaac gccgcctccc acccagatcg 1680
cggtatatag agatatgcat tttattttac ttgtggaaaa atatcggacg acgtggaata 1740
aagagctctt ttcttaaaaa aaaaaaaaaa aaaaaaa 1778

<210> 74
<211> 1422
<212> DNA
<213> Homo sapiens

<400> 74
atgaagaggg cgtccgctgg agggagccgg ctgctggcat gggtgctgtg gctgcaggcc 60

tggcaggtgg cagccccatg cccaggtgcc tgcgtatgct acaatgagcc caaggtgacg 120
acaagctgcc cccagcaggg cctgcaggct gtgcccgtgg gcatccctgc tgccagccag 180
cgcatcttcc tgcacggcaa ccgcatctcg catgtgccag ctgccagctt ccgtgcctgc 240
cgcaacctca ccatcctgtg gctgcactcg aatgtgctgg cccgaattga tgcggctgcc 300
ttcactggcc tggccctcct ggagcagctg gacctcagcg ataatgcaca gctccggtct 360
gtggaccctg ccacattcca cggcctgggc cgcgtacaca cgctgcacct ggaccgctgc 420
ggcctgcagg agctgggccc ggggctgttc cgcggcctgg ctgccctgca gtacctctac 480
ctgcaggaca acgcgctgca ggcactgcct gatgacacct tccgcgacct gggcaacctc 540
acacacctct tcctgcacgg caaccgcatc tccagcgtgc ccgagcgcgc cttccgtggg 600
ctgcacagcc tcgaccgtct cctactgcac cagaaccgcg tggcccatgt gcacccgcat 660
gccttccgtg accttggccg cctcatgaca ctctatctgt ttgccaacaa tctatcagcg 720
ctgcccactg aggccctggc ccccctgcgt gccctgcagt acctgaggct caacgacaac 780
ccctgggtgt gtgactgccg ggcacgccca ctctgggcct ggctgcagaa gttccgcggc 840
tcctcctccg aggtgccctg cagcctcccg caacgcctgg ctggccgtga cctcaaacgc 900
ctagctgcca atgacctgca gggctgcgct gtggccaccg gcccttacca tcccatctgg 960
accggcaggg ccaccgatga ggagccgctg gggcttccca agtgctgcca gccagatgcc 1020
gctgacaagg cctcagtact ggagcctgga agaccagctt cggcaggcaa tgcgctgaag 1080
ggacgcgtgc cgcccggtga cagcccgccg ggcaacggct ctggcccacg gcacatcaat 1140
gactcaccct ttgggactct gcctggctct gctgagcccc cgctcactgc agtgcggccc 1200
gagggctccg agccaccagg gttccccacc tcgggccctc gccggaggcc aggctgttca 1260
cgcaagaacc gcacccgcag ccactgccgt ctgggccagg caggcagcgg gggtggcggg 1320
actggtgact cagaaggctc aggtgcccta cccagcctca cctgcagcct cacccccctg 1380
ggcctggcgc tggtgctgtg gacagtgctt gggccctgct ga 1422

<210> 75
<211> 473
<212> PRT
<213> Homo sapiens

<400> 75
Met Lys Arg Ala Ser Ala Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1 5 10 15

Trp Leu Gln Ala Trp Gln Val Ala Ala Pro Cys Pro Gly Ala Cys Val
20 25 30

Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser Cys Pro Gln Gln Gly Leu
35 40 45

Gln Ala Val Pro Val Gly Ile Pro Ala Ala Ser Gln Arg Ile Phe Leu
50 55 60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Arg Ala Cys
65 70 75 80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Val Leu Ala Arg Ile
85 90 95

Asp Ala Ala Ala Phe Thr Gly Leu Ala Leu Leu Glu Gln Leu Asp Leu
100 105 110

Ser Asp Asn Ala Gln Leu Arg Ser Val Asp Pro Ala Thr Phe His Gly
115 120 125

Leu Gly Arg Val His Thr Leu His Leu Asp Arg Cys Gly Leu Gln Glu
130 135 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr

145 150 155 160

Leu Gln Asp Asn Ala Leu Gln Ala Leu Pro Asp Asp Thr Phe Arg Asp
165 170 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Ser Ser
180 185 190

Val Pro Glu Arg Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
195 200 205

Leu His Gln Asn Arg Val Ala His Val His Pro His Ala Phe Arg Asp
210 215 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Ala
225 230 235 240

Leu Pro Thr Glu Ala Leu Ala Pro Leu Arg Ala Leu Gln Tyr Leu Arg
245 250 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
260 265 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Ser
275 280 285

Leu Pro Gln Arg Leu Ala Gly Arg Asp Leu Lys Arg Leu Ala Ala Asn
290 295 300

Asp Leu Gln Gly Cys Ala Val Ala Thr Gly Pro Tyr His Pro Ile Trp
305 310 315 320

Thr Gly Arg Ala Thr Asp Glu Glu Pro Leu Gly Leu Pro Lys Cys Cys
325 330 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
340 345 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Ser
355 360 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
370 375 380

Gly Thr Leu Pro Gly Ser Ala Glu Pro Pro Leu Thr Ala Val Arg Pro
385 390 395 400

Glu Gly Ser Glu Pro Pro Gly Phe Pro Thr Ser Gly Pro Arg Arg Arg
405 410 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
420 425 430

Gln Ala Gly Ser Gly Gly Gly Gly Thr Gly Asp Ser Glu Gly Ser Gly
435 440 445

Ala Leu Pro Ser Leu Thr Cys Ser Leu Thr Pro Leu Gly Leu Ala Leu
450 455 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465 470

<210> 76
<211> 26
<212> PRT
<213> Homo sapiens

<400> 76
Met Lys Arg Ala Ser Ala Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1 5 10 15

Trp Leu Gln Ala Trp Gln Val Ala Ala Pro
20 25

<210> 77
<211> 447
<212> PRT
<213> Homo sapiens

<400> 77
Cys Pro Gly Ala Cys Val Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser
1 5 10 15

Cys Pro Gln Gln Gly Leu Gln Ala Val Pro Val Gly Ile Pro Ala Ala
20 25 30

Ser Gln Arg Ile Phe Leu His Gly Asn Arg Ile Ser His Val Pro Ala
35 40 45

Ala Ser Phe Arg Ala Cys Arg Asn Leu Thr Ile Leu Trp Leu His Ser
50 55 60

Asn Val Leu Ala Arg Ile Asp Ala Ala Ala Phe Thr Gly Leu Ala Leu
65 70 75 80

Leu Glu Gln Leu Asp Leu Ser Asp Asn Ala Gln Leu Arg Ser Val Asp
85 90 95

Pro Ala Thr Phe His Gly Leu Gly Arg Val His Thr Leu His Leu Asp
100 105 110

Arg Cys Gly Leu Gln Glu Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala
115 120 125

Ala Leu Gln Tyr Leu Tyr Leu Gln Asp Asn Ala Leu Gln Ala Leu Pro
130 135 140

Asp Asp Thr Phe Arg Asp Leu Gly Asn Leu Thr His Leu Phe Leu His
145 150 155 160

Gly Asn Arg Ile Ser Ser Val Pro Glu Arg Ala Phe Arg Gly Leu His
165 170 175

Ser Leu Asp Arg Leu Leu Leu His Gln Asn Arg Val Ala His Val His
180 185 190

Pro His Ala Phe Arg Asp Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe
195 200 205

Ala Asn Asn Leu Ser Ala Leu Pro Thr Glu Ala Leu Ala Pro Leu Arg
210 215 220

Ala Leu Gln Tyr Leu Arg Leu Asn Asp Asn Pro Trp Val Cys Asp Cys
225 230 235 240

Arg Ala Arg Pro Leu Trp Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser
245 250 255

Ser Glu Val Pro Cys Ser Leu Pro Gln Arg Leu Ala Gly Arg Asp Leu
260 265 270

Lys Arg Leu Ala Ala Asn Asp Leu Gln Gly Cys Ala Val Ala Thr Gly
275 280 285

Pro Tyr His Pro Ile Trp Thr Gly Arg Ala Thr Asp Glu Glu Pro Leu
290 295 300

Gly Leu Pro Lys Cys Cys Gln Pro Asp Ala Ala Asp Lys Ala Ser Val
305 310 315 320

Leu Glu Pro Gly Arg Pro Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg
325 330 335

Val Pro Pro Gly Asp Ser Pro Pro Gly Asn Gly Ser Gly Pro Arg His
340 345 350

Ile Asn Asp Ser Pro Phe Gly Thr Leu Pro Gly Ser Ala Glu Pro Pro
355 360 365

Leu Thr Ala Val Arg Pro Glu Gly Ser Glu Pro Pro Gly Phe Pro Thr
370 375 380

Ser Gly Pro Arg Arg Arg Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg
385 390 395 400

Ser His Cys Arg Leu Gly Gln Ala Gly Ser Gly Gly Gly Gly Thr Gly
405 410 415

Asp Ser Glu Gly Ser Gly Ala Leu Pro Ser Leu Thr Cys Ser Leu Thr
420 425 430

Pro Leu Gly Leu Ala Leu Val Leu Trp Thr Val Leu Gly Pro Cys
435 440 445

<210> 78
<211> 20
<212> PRT
<213> Homo sapiens

<400> 78
Gly Ala Leu Pro Ser Leu Thr Cys Ser Leu Thr Pro Leu Gly Leu Ala
1 5 10 15

Leu Val Leu Trp
20

<210> 79
<211> 32
<212> PRT
<213> Homo sapiens

<400> 79
Pro Cys Pro Gly Ala Cys Val Cys Tyr Asn Glu Pro Lys Val Thr Thr
1 5 10 15

Ser Cys Pro Gln Gln Gly Leu Gln Ala Val Pro Val Gly Ile Pro Ala
20 25 30

<210> 80
<211> 24
<212> PRT
<213> Homo sapiens

<400> 80
Ala Ser Gln Arg Ile Phe Leu His Gly Asn Arg Ile Ser His Val Pro
1 5 10 15

Ala Ala Ser Phe Arg Ala Cys Arg
20

<210> 81
<211> 24
<212> PRT
<213> Homo sapiens

<400> 81
Asn Leu Thr Ile Leu Trp Leu His Ser Asn Val Leu Ala Arg Ile Asp
1 5 10 15

Ala Ala Ala Phe Thr Gly Leu Ala
20

<210> 82
<211> 25
<212> PRT
<213> Homo sapiens

<400> 82
Leu Leu Glu Gln Leu Asp Leu Ser Asp Asn Ala Gln Leu Arg Ser Val
1 5 10 15

Asp Pro Ala Thr Phe His Gly Leu Gly
20 25

<210> 83
<211> 24
<212> PRT
<213> Homo sapiens

<400> 83
Arg Val His Thr Leu His Leu Asp Arg Cys Gly Leu Gln Glu Leu Gly
1   5   10   15

Pro Gly Leu Phe Arg Gly Leu Ala
            20

<210> 84
<211> 24
<212> PRT
<213> Homo sapiens

<400> 84
Ala Leu Gln Tyr Leu Tyr Leu Gln Asp Asn Ala Leu Gln Ala Leu Pro
1   5   10   15

Asp Asp Thr Phe Arg Asp Leu Gly
            20

<210> 85
<211> 24
<212> PRT
<213> Homo sapiens

<400> 85
Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Ser Ser Val Pro
1   5   10   15

Glu Arg Ala Phe Arg Gly Leu His
            20

<210> 86
<211> 24
<212> PRT
<213> Homo sapiens

<400> 86
Ser Leu Asp Arg Leu Leu Leu His Gln Asn Arg Val Ala His Val His
1   5   10   15

Pro His Ala Phe Arg Asp Leu Gly
            20

<210> 87
<211> 24
<212> PRT
<213> Homo sapiens

<400> 87

Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Ala Leu Pro
1 5 10 15

Thr Glu Ala Leu Ala Pro Leu Arg
20

<210> 88
<211> 51
<212> PRT
<213> Homo sapiens

<400> 88
Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp Ala Trp Leu
1 5 10 15

Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Ser Leu Pro Gln
20 25 30

Arg Leu Ala Gly Arg Asp Leu Lys Arg Leu Ala Ala Asn Asp Leu Gln
35 40 45

Gly Cys Ala
50

<210> 89
<211> 421
<212> PRT
<213> Homo sapiens

<400> 89
Cys Pro Gly Ala Cys Val Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser
1 5 10 15

Cys Pro Gln Gln Gly Leu Gln Ala Val Pro Val Gly Ile Pro Ala Ala
20 25 30

Ser Gln Arg Ile Phe Leu His Gly Asn Arg Ile Ser His Val Pro Ala
35 40 45

Ala Ser Phe Arg Ala Cys Arg Asn Leu Thr Ile Leu Trp Leu His Ser
50 55 60

Asn Val Leu Ala Arg Ile Asp Ala Ala Ala Phe Thr Gly Leu Ala Leu
65 70 75 80

Leu Glu Gln Leu Asp Leu Ser Asp Asn Ala Gln Leu Arg Ser Val Asp
85 90 95

Pro Ala Thr Phe His Gly Leu Gly Arg Val His Thr Leu His Leu Asp
100 105 110

Arg Cys Gly Leu Gln Glu Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala
115 120 125

Ala Leu Gln Tyr Leu Tyr Leu Gln Asp Asn Ala Leu Gln Ala Leu Pro
130 135 140

Asp Asp Thr Phe Arg Asp Leu Gly Asn Leu Thr His Leu Phe Leu His
145 150 155 160

Gly Asn Arg Ile Ser Ser Val Pro Glu Arg Ala Phe Arg Gly Leu His
165 170 175

Ser Leu Asp Arg Leu Leu Leu His Gln Asn Arg Val Ala His Val His
180 185 190

Pro His Ala Phe Arg Asp Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe
195 200 205

Ala Asn Asn Leu Ser Ala Leu Pro Thr Glu Ala Leu Ala Pro Leu Arg
210 215 220

Ala Leu Gln Tyr Leu Arg Leu Asn Asp Asn Pro Trp Val Cys Asp Cys
225 230 235 240

Arg Ala Arg Pro Leu Trp Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser
245 250 255

Ser Glu Val Pro Cys Ser Leu Pro Gln Arg Leu Ala Gly Arg Asp Leu
260 265 270

Lys Arg Leu Ala Ala Asn Asp Leu Gln Gly Cys Ala Val Ala Thr Gly
275 280 285

Pro Tyr His Pro Ile Trp Thr Gly Arg Ala Thr Asp Glu Glu Pro Leu
290 295 300

Gly Leu Pro Lys Cys Cys Gln Pro Asp Ala Ala Asp Lys Ala Ser Val
305 310 315 320

Leu Glu Pro Gly Arg Pro Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg
325 330 335

Val Pro Pro Gly Asp Ser Pro Pro Gly Asn Gly Ser Gly Pro Arg His
340 345 350

Ile Asn Asp Ser Pro Phe Gly Thr Leu Pro Gly Ser Ala Glu Pro Pro
355 360 365

Leu Thr Ala Val Arg Pro Glu Gly Ser Glu Pro Pro Gly Phe Pro Thr
370 375 380

Ser Gly Pro Arg Arg Arg Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg
385 390 395 400

Ser His Cys Arg Leu Gly Gln Ala Gly Ser Gly Gly Gly Gly Thr Gly
405 410 415

Asp Ser Glu Gly Ser
420

<210> 90
<211> 5094
<212> DNA

<213> Homo sapiens

<400> 90
gcgaaacggc agaggagccg agccccctcc gcccaaggcg ccctccctcc gtccgcgcac 60
aggcgccgtc gcttggagga gcaaggtgcc tcccagcccg caggggcgcc gcgcgcaagc 120
ccgcgggctc ttcggtggct ctgccccggg actgcacctg gaggcggccc cggacgggga 180
tggtcagcgg ctgctgccgt ctggctcgcg agcgggacgc tgtgagggca ccatggcgct 240
gactcccggg tgggggtcct cggcggggcc ggtccggccg gagctctggc tgctgctgtg 300
ggcagccgcg tggcgcctgg gtgcctcggc gtgccccgcc ctctgcacct gcaccggaac 360
cacggtggac tgccacggca cggggctgca ggccattccc aagaatatac ctcggaacac 420
cgagcgcctg gaactcaatg gcaacaacat cactcggatc cataagaatg actttgcggg 480
gctcaagcag ctgcgggtgc tgcagctgat ggagaaccag attggagcag tggaacgtgg 540
tgcttttgat gacatgaagg agctggagcg gctgcgactg aaccgaaacc agctgcacat 600
gttaccggaa ctgctgttcc agaacaacca ggctttgtca agactggact tgagtgagaa 660
cgccatccag gccatcccca ggaaagcttt tcggggagct acggacctta aaaatttacg 720
gctggacaag aaccagatca gctgcattga ggaaggggcc ttccgtgctc tgcggggct 780
ggaggtgctg accctgaaca acaacaatat caccaccatc cccgtgtcca gcttcaacca 840
tatgcccaag ctacggacct tccgcctgca ctccaaccac ctgttttgcg actgccacct 900
ggcctggctc tcgcagtggc tgaggcagcg gccaaccatc gggctcttca cccagtgctc 960
gggcccagcc agcctgcgtg gcctcaatgt ggcagaggtc cagaagagtg agttcagctg 1020
ctcaggccag ggagaagcgg ggcgcgtgcc cacctgcacc ctgtcctccg gctcctgccc 1080
ggccatgtgc acctgcagca atggcatcgt ggactgtcgt ggaaaaggcc tcactgccat 1140
cccggccaac ctgcccgaga ccatgacgga gatacgcctg gagctgaacg gcatcaagtc 1200
catccctcct ggagccttct caccctacag aaagctacgg aggatagacc tgagcaacaa 1260
tcagatcgct gagattgcac ccgacgcctt ccagggcctc cgctccctga actcgctggt 1320
cctctatgga aacaagatca cagacctccc ccgtggtgtg tttggaggcc tatacaccct 1380
acagctcctg ctcctgaatg ccaacaagat caactgcatc cggcccgatg ccttccagga 1440
cctgcagaac ctctcactgc tctccctgta tgacaacaag atccagagcc tcgccaaggg 1500
cactttcacc tccctgcggg ccatccagac tctgcacctg gcccagaacc ctttcatttg 1560
cgactgtaac ctcaagtggc tggcagactt cctgcgcacc aatcccatcg agacgagtgg 1620
tgcccgctgt gccagtcccc ggcgcctcgc caacaagcgc atcgggcaga tcaagagcaa 1680
gaagttccgg tgctcagcca aagagcagta cttcattcca ggcacggagg attaccagct 1740
gaacagcgag tgcaacagcg acgtggtctg tccccacaag tgccgctgtg aggccaacgt 1800
ggtggagtgc tccagcctga agctcaccaa gatccctgag cgcatccccc agtccacggc 1860
agaactgcga ttgaataaca atgagatttc catcctggag gccactggga tgtttaaaaa 1920
acttacacat ctgaagaaaa tcaatctgag caacaacaag gtgtcagaaa ttgaagatgg 1980
ggccttcgag ggcgcagcct ctgtgagcga gctgcaccta actgccaacc agctggagtc 2040
catccggagc ggcatgttcc ggggtctgga tggcttgagg accctaatgc tgcggaacaa 2100
ccgcatcagc tgcatccaca acgacagctt cacgggcctg cgcaacgtcc ggctcctctc 2160
gctctacgac aaccagatca ccaccgtatc cccaggagcc ttcgacaccc tccagtccct 2220
ctccacactg aatctcctgg ccaacccttt caactgcaac tgccagctgg cctggctagg 2280
aggctggcta cggaagcgca agatcgtgac ggggaacccg cgatgccaga accctgactt 2340
tttgcggcag attcccctgc aggacgtggc cttccctgac ttcaggtgtg aggaaggcca 2400
ggaggagggg ggctgcctgc cccgcccaca gtgcccacag gagtgcgcct gcctggacac 2460
cgtggtccga tgcagcaaca agcacctgcg ggccctgccc aagggcattc ccaagaatgt 2520
cacagaactc tatttggacg ggaaccagtt cacgctggtt ccgggacagc tgtctacctt 2580
caagtacctg cagctcgtgg acctgagcaa caacaagatc agttccttaa gcaattcctc 2640
cttcaccaac atgagccagc tgaccactct gatcctcagc tacaatgccc tgcagtgcat 2700
cccgcctttg gccttccagg gactccgctc cctgcgcctg ctgtctctcc acggcaatga 2760
catctccacc ctccaagagg gcatctttgc agacgtgacc tccctgtctc acctggccat 2820
tggtgccaac cccctatact gtgactgcca cctccgctgg ctgtccagct gggtgaagac 2880
tggctacaag gaaccgggca ttgctcgttg tgctgggccc caggacatgg agggcaagct 2940
gctcctcacc acgcctgcca agaagtttga atgccaaggt cctccaacgc tggctgtcca 3000
ggccaagtgt gatctctgct tgtccagtcc gtgccagaac cagggcacct gccacaacga 3060
ccccccttgag gtgtacaggt gcgcctgccc cagcggctat aagggtcgag actgtgaggt 3120
gtccctgaac agctgttcca gtggcccctg tgaaaatggg ggcacctgcc atgcacagga 3180
gggcgaggat gccccgttca cgtgctcctg tcccaccggc tttgaaggac caacctgtgg 3240
ggtgaacaca gatgactgtg tggatcatgc ctgtgccaat gggggcgtct gtgtggatgg 3300

tgtgggcaac tacacctgcc agtgccccct gcagtatgag ggaaaggcct gtgagcagct 3360
ggtggacttg tgctctccgg atctgaaccc atgtcaacac gaggcccagt gtgtgggcac 3420
cccggatggg cccaggtgtg agtgcatgcc aggttatgca ggtgacaact gcagtgagaa 3480
ccaggatgac tgcagggacc accgctgcca gaatggggcc cagtgtatgg atgaagtcaa 3540
cagctactcc tgcctctgtg ctgagggcta cagtggacag ctctgtgaga tccctcccca 3600
tctgcctgcc cccaagagcc cctgtgaggg gactgagtgc cagaatgggg ccaactgtgt 3660
ggaccagggc aacaggcctg tgtgccagtg cctcccaggc ttcggtggcc ctgagtgtga 3720
gaagttgctc agtgtcaact ttgtggatcg ggacacttac ctgcagttca ctgacctgca 3780
aaactggcca cgggccaaca tcacgttgca ggtctccacg gcagaggaca atgggatcct 3840
tctgtacaac ggggacaacg accacattgc agttgagctg taccagggcc atgtgcgtgt 3900
cagctacgac ccaggcagct accccagctc tgccatctac agtgctgaga cgatcaacga 3960
tgggcaattc cacaccgttg agctggttgc ctttgaccag atggtgaatc tctccattga 4020
tggcgggagc ccatgacca tggacaactt tggcaaacat tacacgctca acagcgaggc 4080
gccactctat gtgggaggga tgcccgtgga tgtcaactca gctgccttcc gcctgtggca 4140
gatcctcaac ggcaccggct tccacggttg catccgaaac ctgtacatca acaacgagct 4200
gcaggacttc accaagacgc agatgaagcc aggcgtggtg ccaggctgcg aaccctgccg 4260
caagctctac tgcctgcatg gcatctgcca gcccaatgcc accccagggc ccatgtgcca 4320
ctgcgaggct ggctgggtgg gcctgcactg tgaccagccc gctgacggcc cctgccatgg 4380
ccacaagtgt gtccatgggc aatgcgtgcc cctcgacgct ctttcctaca gctgccagtg 4440
ccaggatggg tactcggggg cactgtgcaa ccaggccggg gccctggcag agccctgcag 4500
aggcctgcag tgcctgcatg gccactgcca ggcctcaggc accaaggggg cacactgtgt 4560
gtgtgacccc ggcttttcgg gcgagctgtg tgagcaagag tccgagtgcc gggggaccc 4620
tgtccgggac tttcaccagg tccagagggg ctatgccatc tgccagacca cgcgccccct 4680
gtcatgggtg gagtgccggg gctcgtgccc aggccagggc tgctgccagg gccttcggct 4740
gaagcggagg aagttcacct ttgagtgcag cgatgggacc tcttttgccg aggaggtgga 4800
aaagcccacc aagtgtggct gtgccctctg cgcatagcgc tgggcgtgga caggccggtg 4860
agggcgggca aggggcccca gccgctgcag cagcggagac agtcgccagc agctgggctg 4920
gggtgcaggt catcacagga cggctcctgg gcagctgggc cctcctgggt ggggtggtgc 4980
cagagcagcc ttttaaaagc aaattgcgcc atagctgggg gcagcggggg tgggcgaggc 5040
ctgagctgcg ggctgccctc tccggaagtc cttgcacaaa taggcgctta ataa 5094

<210> 91
<211> 557
<212> DNA
<213> Homo sapiens

<400> 91
agtagagaca gggtttcacc atgttagcca agatggtctc gatctcctga cctcgtgatc 60
cgcccacctg ggcctcccaa agtgctggga ttacaggcgt gagcccccgc acccggccaa 120
aggcaggatt atttgttaag cgggttgcag tgttctacca gaccccccac cctcaaccca 180
agcccctggt ctcactggga cttggccact gccaggagcg tctggccctt cacggcttca 240
ggcccagcac agcgcgtgtc attctgggaa aacatcttgt ctttttgggc ctgaagccaa 300
cgatagaggt cgctcaggtt ctggtcacag atccaggggt tgccggagat gtcgaagcca 360
tcccgcatgt cccagtttgg ctgccctagg gatgcccana gcccctcggg cacgctggcc 420
agtgagttat tggagaggtc cagcatgtnc agctgccgca ggccctggaa ggcaccggct 480
gccaccctgg ccagcttgtt gccgttcaag aagaggtagc gcangtcctg ctgcggcaag 540
aggagatctt ttcccag 557

<210> 92
<211> 7452
<212> DNA
<213> Homo sapiens

<400> 92
tgatcggaac tgaaagacct cccgcgatac ctggcagagg cagtggctct tccctgtggt 60
ccagggctga ctgactttga aggtaatttc agtcaaccca gcctttactg ggctctgact 120
gcattaggct gcatcaaagg ggattggatc ccatgattct ttatatcttc tgacattaag 180
cctttgtcag ctataggtgt tacaaatatc tttagtttgt ggtttatctt ttcccctttt 240

ttatggtgtc ttgaaggata gaagtcttaa tgcagacagc attatcagtg tgttcaaaag 300
acagctagac acgttttgcc tatagacaaa tgggcaaaag gaaacccagc tttctcaaat 360
gaagcacaag tgggccttaa ttatgtgaaa aggtgttcaa gttcatcatt aaacagggaa 420
aggaaaagtt aaaaccatgc tgagatatct ttcatagaaa tggcaaaaag caggaagtgc 480
cacgtgtggg cagagaggaa gcacaggaac tctcacaaat ggcaggtgtc atcgtagacc 540
aacacaacca ctttggagag cagtttgact ttccccagtt aaactgaaca tgtgagcggc 600
cgggcgtggt ggctcatgcc tgtaatccca gcagtttggg aggccgaggc gggcggattg 660
cctgagctca ggagttcaag accagccagg gcaacacggt aaaaccccgt ctctactaaa 720
atacaaaaaa ttagctgggc gtgatggtgt gtgcctgtaa tcccagctac ttgtgaggcc 780
gaggcaggag aattgcttga accagggagc aggaggttgc agtgagccga gatcgcacca 840
ctgcacccca gcctggcgac agagtccccc tcccccacca aaaaaacaac aagtgagcat 900
cctgcaacct agcaatgcca ttgttgaaca agttcaaaga tgttcttagc cttattagtc 960
ccaaaaggaa gaaaaaaatg gaggatttga gaatgttctt agctttattg ctaagcggag 1020
aaagaaaaac aacacatacc aaaaaaaaaa aaaaaaaaaa aaaaaaacaa aaaacctggg 1080
tgggaaatta gggccatgtg gcatgaaaag gaagacccag gggaagtgtg gcccatctag 1140
gggtgtggct actgcagtga tccagctgta tcactgaact tccctggcat catagagtta 1200
tattgtgcca tttatggaaa aactctcccc actgctcttg gctttgacag taggaatcag 1260
gttatatatg gtctctcggt ttgaagatat ttgtcattaa aaaccagaac aagggctctg 1320
agatagggtc ctttcctgac ctactctggt aaagtcttta tcctcaggat gcaaggatac 1380
caccctcttc ctgtggaaag tgtcgaatca catgcagagc tctaagtctt tcagttactt 1440
tggagtgcag aaccatttca ggtaaggcca aatattttaa acattagtat aggaaattag 1500
agggctcttt agtctgtgtg tgcatgagaa gtaaaattgc acgagaagca atttatgtaa 1560
aatttcgctt aggaaacatt gttttggcag gttagtagta tggtgtgcat ttcccagaaa 1620
attcagtgcc gtgagtatta cctttagtta agcatcttag aaatagtagc tctttatgtt 1680
tatggctaag tcagaaatac taccctcaaa ttctatgtga ccctagttat actgttgagc 1740
ctttctgtgc ctctgtgcct tcatccttga atcggggata atatacttac ctcctaaggt 1800
tattgtaagg attaaatgca tgtagtataa ataaagagct gagaacaatg catggcgtaa 1860
agtgataggt attattatat gcttttgttg gctgttgatt gaaggtgttc gctgttttgg 1920
gggtgtcctt taatagagta acttggtact gtggaaatag catgattgtg agcaaaagaa 1980
tcagatggtg gtggctgcag actttgctgt tcccttcttg actgttggtt atagccaatg 2040
cagggtaagt tataaagtca agagcagagc cgttttcaca atggacattg ctttgtgatg 2100
tctgtgagct tgaatgtgag aatgattatt ttaattctct atgtaaagac tttaaagtat 2160
tggctattcg gtagcttgaa aactctgtaa tctcatgctt taaactgaga gtggaaaatc 2220
aataaagcaa aagcatgagg ccacgcagtg tagaatgagt gtcttttcac cacgtaggga 2280
aatctgtagt cctaagaaaa gagggagtga gaattctggc gaaaagattg tgcctctgca 2340
caaagtgcag gatcccaggg ttcagtacag gcgcgaacgc tcctgtgtgt tgaccacact 2400
cccacggttg ctttttcaga catgctgagg gggactctac tgtgcgcggt gctcgggctt 2460
ctgcgcgccc agcccttccc ctgtccgcca gcttgcaagt gtgtcttccg ggacgccgcg 2520
cagtgctcgg ggggcgacgt ggcgcgcatc tccgcgctgg gcctgcccac caacctcacg 2580
cacatcctgc tcttcggaat gggccgcggc gtcctgcaga gccagagctt cagcggcatg 2640
accgtcctgc agcgcctcat gatctccgac agccacattt ccgccgttgc cccggcacc 2700
ttcagtgacc tgataaaact gaaaaccctg aggctgtcgc gcaacaaaat cacgcatctt 2760
ccaggtgcgc tgctggataa gatggtgctc ctggagcagt tgtttttgga ccacaatgcg 2820
ctaaggggca ttgaccaaaa catgtttcag aaactggtta acctgcagga gctcgctctg 2880
aaccagaatc agctcgattt ccttcctgcc agtctcttca cgaatctgga gaacctgaag 2940
ttgttggatt tatcgggaaa caacctgacc cacctgccca aggggttgct tggagcacag 3000
gctaagctcg agagacttct gctccactcg aaccgccttg tgtctctgga ttcggggctg 3060
ttgaacagcc tgggcgccct gacggagctg cagttccacc gaaatcacat ccgttccatc 3120
gcacccgggg ccttcgaccg gctcccaaac ctcagttctt tgacgctttc gagaaaccac 3180
cttgcgtttc tcccctctgc gctctttctt cattcgcaca atctgactct gttgactctg 3240
ttcgagaacc cgctggcaga gctcccgggg gtgctcttcg ggagatggg gggcctgcag 3300
gagctgtggc tgaaccgcac ccagctgcgc accctgcccg ccgccgcctt ccgcaacctg 3360
agccgcctgc ggtacttagg ggtgactctg agcccgcggc tgagcgcgct tccgcagggc 3420
gccttccagg gccttggcga gctccaggtg ctcgccctgc actccaacgg cctgaccgcc 3480
ctccccgacg gcttgctgcg cggcctcggc aagctgcgcc aggtgtccct gcgccgcaac 3540
aggctgcgcg ccctgccccg tgccctcttc cgcaatctca gcagcctgga gagcgtccag 3600
ctcgaccaca accagctgga gaccctgcct ggcgacgtgt ttggggctct gccccggctg 3660
acggaggtcc tgttggggca caactcctgg cgctgcgact gtggcctggg gcccttcctg 3720

gggtggctgc ggcagcacct aggcctcgtg ggcggggaag agcccccacg gtgcgcaggc 3780
cctggggcgc acgccggcct gccgctctgg gccctgccgg ggggtgacgc ggagtgcccg 3840
ggcccccggg gcccgcctcc ccgccccgct gcggacagct cctcggaagc ccctgtccac 3900
ccagccttgg ctcccaacag ctcagaaccc tgggtgtggg cccagccggt gaccacgggc 3960
aaaggtcaag atcatagtcc gttctggggg ttttattttc tgcttttagc tgttcaggcc 4020
atgatcaccg tgatcatcgt gtttgctatg attaaaattg gccaactctt tcgaaaatta 4080
atcagagaga gagcccttgg gtaaaccaat gggaaaatct tctaattact tagaacctga 4140
ccagatgtgg ctcggagggg aatccagacc cgctgctgtc ttgctctccc tcccctcccc 4200
actcctcctc tcttcttcct cttctctctc actgccacgc cttcctttcc ctcctcctcc 4260
ccctctccgc tctgtgctct tcattctcac gggcccgcaa cccctcctct ctctgtcccc 4320
gcccgtctct ggaaactgag cttgacgttt gtaaactgtg gttgcctgcc ttcccagctc 4380
cacgcggtgt gcgctgacac tgccgggggg ctggactgtg ttggacccat ccttgccccg 4440
ctgtgcctgg cttggcctct ggtggagaga gggacctctt cagtgtctac tgagtaaggg 4500
gacagctcca ggccggggct gtctcctgca cagagtaagc cggtaaatgt ttgtgaaatc 4560
aatgcgtgga taaaggaaca catgccatcc aagtgatgat ggcttttcct ggagggaaag 4620
gataggctgt tgctctatct aatttttgt ttttgtttt ggacagtcta gctctgtggc 4680
ccaggctggc gtgcagtggg ccgtctcagt tcactgcagc ctccgccctc caggttcaag 4740
tgattctcat gcctcagcgt tctgagtagc tgggattaga ggcgtgtgcc actacacccg 4800
gctaattttt gtacttttta aagtagagac gggctttgcc atattggcct ggctgatctc 4860
aaactcctgg tcttgaactc ctggccacaa gtgatctgcc cgccttagcc tcccaaagtg 4920
ctgggattac aggcgcaagc cactacacct gccctcttca tcgaatttta tttgagaagt 4980
agagctcttg ccattttttc ccttgctcca tttttctcac tttatgtctc tctgacctat 5040
gggctacttg ggagagcact ggactccatt catgcatgag cattttcagg ataagcgact 5100
tctgtgaggc tgagagagga agaaaacacg gagccttccc tccaggtgcc cagtgtaggt 5160
ccagcgtgtt tcctgagcct cctgtgagtt tccacttgct ttacatccat gcaacatgtc 5220
attttgaaac tggattgatt tgcatttcct ggaactctgc cacctcattt cacaagcatt 5280
tatggagcag ttaacatgtg actggtattc atgaatataa tgataagctt gattctagtt 5340
cagctgctgt cacagtctca tttgttcttc caactgaaag ccgtaaaacc tttgttgctt 5400
taattgaatg tctgtgctta tgagaggcag tggttaaaac attttctggc gagttgacaa 5460
ctgtgggttc aaatcccagc tctaccactt actaactgca tgggactttg ggtaagacac 5520
ctgcttacat tctctaagcc ttggtttcct gaaccttaaa acaggataac atagtacctg 5580
cttcatagag ttttgtgaga attaaaggca ataaagcata taatgactta gcccagcggc 5640
ctgcagacaa tacatgttaa tgaatgttag ctattattac taaagatgag caattattat 5700
tggcatcatg atttctaaag aagagctttg agttggtatt tttctctgtg tataagggta 5760
agtccgaact ttctcatact ggaggttaca ttcacatcag tctgtcttcc cctgcggatg 5820
gcctcagccc tgggtggcca ggctctgtgc tcacagtcca gagcaatgga tcctccaaca 5880
ccaccaggtg gatgtggagc aggagagctg gatcgtggca tttgtttctg ggttctgcag 5940
ttgggagttg gtttctgggt tctccattgg tctacttgtc tagtcccata ccagactcac 6000
ggtctccatt attggagctt taataatttt tggtataggg tcatctctcc accttgtttt 6060
tcttctattc ttggttcttt gcaattctat gaatatttca gggtcagcat gtcaactcca 6120
ttgaaaaacc ctgctgggat tttaatagaa cttacagctc acgcctgtaa tcccagcact 6180
ttgggaggct gaggtgggtg gatcacaggt caggagtttg agaacagctg gccaagatgg 6240
tgaaaccccg tctctactaa aaatacaaaa attagctggg tgcggtggca ggtgcctgta 6300
gtcccagcta cttgggacac cgaggcagga gaatcacttg aacccgggag gcggaggttg 6360
cagtgagccg agatcgtgcc actgcactct agcctgggcg acagagcgag actccatctc 6420
aaaaaaaaag aaaaagaaaa ttgcagtaaa tttaaaacta atttggggaa gaatctgtat 6480
ttttacaata cctagtgttc ttgccagtaa gcatggttca tcttcccatt tatttacgtc 6540
attttaaatc tttcagtgat gttttagaat ttttttata aaaaccttca ctataagaac 6600
agaaaaccaa acaccgcatg ttctcactca taggtgggaa ttgaacaatg agaacacttg 6660
gacacagggc ggggaacgtc acacgcctgg actgttgggg gggtggctgg gagagggata 6720
gtgttaggag aaatacctaa tgtaaatgac gagttaatgg tgcagccaac caacctggca 6780
catgtattca tatgtaacaa acctgcacgt tgtgcacatg taccctagaa cttaaagtat 6840
attaaaaaaa gaaaccttgg cactgatttt gttagattta ttcctaggta tccttcctct 6900
tttttgattt gtcattgcta ttgtagatgg catcttttta aaaagttata ttttctaaag 6960
caaaaaataa aaaaagttgt atttctaatt tttattacca atatataaga atgtaattta 7020
tttttacata attatcttat gtctagtaat aattctgata atttgcttct tcctattaaa 7080
accttacacc cattattgat ttatttttct gttttaaaat atcttcctgc actggctaaa 7140
acctccacta taatgttgag cagaacagtg aggcatcctt agaactatct tggttgcaaa 7200

gggtaggtct ctaatgtttc atcaataaat gtgatgtttc tagtctgagt ttgctaagta 7260
tattttaaaa taatcagtaa agttagattt tatccatttt tatcttaact attgagatgc 7320
tcatatcatt tttcttcttc aatgtgttaa aatggtgaat aaatttatag attttggaaa 7380
agtaaattca ttcttgcatt cccgaagtaa accaagccat gctatgtgta tttaaaatat 7440
attgctgaat tc 7452

<210> 93
<211> 1946
<212> DNA
<213> Homo sapiens

<400> 93
cgcgctgcga gcgccccgcc agtccgcgcc gccgccctca ccctgtgcgc ccgcagcccg 60
cgagcccagc ccggccccggt agagcggagc gccggagcct cgtcccgcgg ccgggccggg 120
accgggccgg agcagcggcg cctggatgcg gacccggccg cgcgcagacg ggcgcccgcc 180
ccgaagccgc ttccagtgcc cgacgcgccc cgctcgaccc cgaagatgaa gagggcgtcc 240
tccggaggaa gcaggctgct ggcatgggtg ttatggctac aggcctggag ggtagcaaca 300
ccatgccctg gtgcttgtgt gtgctacaat gagcccaagg taacaacaag ctgcccccag 360
cagggtctgc aggctgtgcc cactggcatc ccagcctcta gccagcgaat cttcctgcat 420
ggcaaccgaa tctctcacgt gccagctgcg agcttccagt catgccgaaa tctcactatc 480
ctgtggctgc actctaatgc gctggctcgg atcgatgctg ctgccttcac tggtctgacc 540
ctcctggagc aactagatct tagtgataat gcacagcttc atgtcgtgga ccctaccacg 600
ttccacggcc tgggccacct gcacacactg cacctagacc gatgtggcct gcgggagctg 660
ggtcccggcc tattccgtgg actagcagct ctgcagtacc tctacctaca agacaacaat 720
ctgcaggcac tccctgacaa cacctttcga gacctgggca acctcacgca tctctttctg 780
catggcaacc gtatccccag tgtgcctgag cacgctttcc gtggcctgca cagtcttgac 840
cgcctcctct tgcaccagaa ccatgtggct cgtgtgcacc cacatgcctt ccgggacctt 900
ggccgcctca tgaccctcta cctgtttgcc aacaacctct ccatgctgcc tgcagaggtc 960
ctaatgcccc tgaggtctct gcagtacctg cgactcaatg acaacccctg ggtgtgtgac 1020
tgccgggcac gtccactctg ggcctggctg cagaagttcc gaggttcctc atcagaggtg 1080
ccctgcaacc tgccccaacg cctggcagac cgtgatctta agcgcctcgc tgccagtgac 1140
ctagagggct gtgctgtggc ttcaggaccc ttccgtccca tccagaccag tcagctcact 1200
gatgaggagc tgctgagcct ccccaagtgc tgccagccag atgctgcaga caaagcctca 1260
gtactggaac ccgggaggcc agcttctgcc ggaaacgccc tcaagggacg tgtgcctccc 1320
ggtgacactc caccaggcaa tggctcaggc cctcggcaca tcaatgactc tccatttgga 1380
actttgccca gctctgcaga gcccccactg actgccctgc ggcctggggg ttccgagcca 1440
ccaggacttc ccaccactgg tccccgcagg aggccaggtt gttcccggaa gaatcgcacc 1500
cgcagccact gccgtctggg ccaggcggga agtggggcca gtggaacagg ggacgcagag 1560
ggttcagggg ctctgcctgc tctggcctgc agccttgctc ctctgggcct tgcactggta 1620
ctttggacag tgcttgggcc ctgctgacca gccaccagcc accaggtgtg tgtacatatg 1680
gggtctccct ccacgccgcc agccagagcc agggacaggc tctgaggggc aggccaggcc 1740
ctccctgaca gatgcctccc caccagccca cccccatctc caccccatca tgtttacagg 1800
gttccggggg tggcgtttgt tccagaacgc cacctcccac ccggatcgcg gtatatagag 1860
atatgaattt tattttactt gtgtaaaata tcggatgacg tggaataaag agctcttttc 1920
ttaaaaaaaa aaaaaaaaaa aaaaaa 1946

<210> 94
<211> 1422
<212> DNA
<213> Mus musculus

<400> 94
atgaagaggg cgtcctccgg aggaagcagg ctgctggcat gggtgttatg gctacaggcc 60
tggagggtag caacaccatg ccctggtgct tgtgtgtgct acaatgagcc caaggtaaca 120
acaagctgcc cccagcaggg tctgcaggct gtgcccactg gcatcccagc ctctagccag 180
cgaatcttcc tgcatggcaa ccgaatctct cacgtgccag ctgcgagctt ccagtcatgc 240
cgaaatctca ctatcctgtg gctgcactct aatgcgctgg ctcggatcga tgctgctgcc 300
ttcactggtc tgaccctcct ggagcaacta gatcttagtg ataatgcaca gcttcatgtc 360

```
gtggacccta ccacgttcca cggcctgggc cacctgcaca cactgcacct agaccgatgt 420
ggcctgcggg agctgggtcc cggcctattc cgtggactag cagctctgca gtacctctac 480
ctacaagaca acaatctgca ggcactccct gacaacacct ttcgagacct gggcaacctc 540
acgcatctct ttctgcatgg caaccgtatc cccagtgtgc ctgagcacgc tttccgtggc 600
ctgcacagtc ttgaccgcct cctcttgcac cagaaccatg tggctcgtgt gcacccacat 660
gccttccggg accttggccg cctcatgacc ctctacctgt ttgccaacaa cctctccatg 720
ctgcctgcag aggtcctaat gcccctgagg tctctgcagt acctgcgact caatgacaac 780
ccctgggtgt gtgactgccg ggcacgtcca ctctgggcct ggctgcagaa gttccgaggt 840
tcctcatcag aggtgccctg caacctgccc caacgcctgg cagaccgtga tcttaagcgc 900
ctcgctgcca gtgacctaga gggctgtgct gtggcttcag gacccttccg tcccatccag 960
accagtcagc tcactgatga ggagctgctg agcctcccca agtgctgcca gccagatgct 1020
gcagacaaag cctcagtact ggaacccggg aggccagctt ctgccggaaa cgccctcaag 1080
ggacgtgtgc ctcccggtga cactccacca ggcaatggct caggccctcg gcacatcaat 1140
gactctccat ttggaacttt gcccagctct gcagagcccc cactgactgc cctgcggcct 1200
gggggttccg agccaccagg acttcccacc actggtcccc gcaggaggcc aggttgttcc 1260
cggaagaatc gcacccgcag ccactgccgt ctgggccagg cgggaagtgg ggccagtgga 1320
acaggggacg cagagggttc aggggctctg cctgctctgg cctgcagcct tgctcctctg 1380
ggccttgcac tggtactttg gacagtgctt gggccctgct ga                    1422

<210> 95
<211> 473
<212> PRT
<213> Mus musculus

<400> 95
Met Lys Arg Ala Ser Ser Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1               5                   10                  15

Trp Leu Gln Ala Trp Arg Val Ala Thr Pro Cys Pro Gly Ala Cys Val
            20                  25                  30

Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser Cys Pro Gln Gln Gly Leu
        35                  40                  45

Gln Ala Val Pro Thr Gly Ile Pro Ala Ser Ser Gln Arg Ile Phe Leu
    50                  55                  60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Gln Ser Cys
65                  70                  75                  80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Ala Leu Ala Arg Ile
                85                  90                  95

Asp Ala Ala Ala Phe Thr Gly Leu Thr Leu Leu Glu Gln Leu Asp Leu
            100                 105                 110

Ser Asp Asn Ala Gln Leu His Val Val Asp Pro Thr Thr Phe His Gly
        115                 120                 125

Leu Gly His Leu His Thr Leu His Leu Asp Arg Cys Gly Leu Arg Glu
    130                 135                 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145                 150                 155                 160

Leu Gln Asp Asn Asn Leu Gln Ala Leu Pro Asp Asn Thr Phe Arg Asp
                165                 170                 175
```

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Pro Ser
180 185 190

Val Pro Glu His Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
195 200 205

Leu His Gln Asn His Val Ala Arg Val His Pro His Ala Phe Arg Asp
210 215 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Met
225 230 235 240

Leu Pro Ala Glu Val Leu Met Pro Leu Arg Ser Leu Gln Tyr Leu Arg
245 250 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
260 265 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Asn
275 280 285

Leu Pro Gln Arg Leu Ala Asp Arg Asp Leu Lys Arg Leu Ala Ala Ser
290 295 300

Asp Leu Glu Gly Cys Ala Val Ala Ser Gly Pro Phe Arg Pro Ile Gln
305 310 315 320

Thr Ser Gln Leu Thr Asp Glu Glu Leu Leu Ser Leu Pro Lys Cys Cys
325 330 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
340 345 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Thr
355 360 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
370 375 380

Gly Thr Leu Pro Ser Ser Ala Glu Pro Pro Leu Thr Ala Leu Arg Pro
385 390 395 400

Gly Gly Ser Glu Pro Pro Gly Leu Pro Thr Thr Gly Pro Arg Arg Arg
405 410 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
420 425 430

Gln Ala Gly Ser Gly Ala Ser Gly Thr Gly Asp Ala Glu Gly Ser Gly
435 440 445

Ala Leu Pro Ala Leu Ala Cys Ser Leu Ala Pro Leu Gly Leu Ala Leu
450 455 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465 470

<210> 96
<211> 26
<212> PRT
<213> Mus musculus

<400> 96
Met Lys Arg Ala Ser Ser Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1 5 10 15

Trp Leu Gln Ala Trp Arg Val Ala Thr Pro
20 25

<210> 97
<211> 447
<212> PRT
<213> Homo sapiens

<400> 97
Cys Pro Gly Ala Cys Val Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser
1 5 10 15

Cys Pro Gln Gln Gly Leu Gln Ala Val Pro Thr Gly Ile Pro Ala Ser
20 25 30

Ser Gln Arg Ile Phe Leu His Gly Asn Arg Ile Ser His Val Pro Ala
35 40 45

Ala Ser Phe Gln Ser Cys Arg Asn Leu Thr Ile Leu Trp Leu His Ser
50 55 60

Asn Ala Leu Ala Arg Ile Asp Ala Ala Ala Phe Thr Gly Leu Thr Leu
65 70 75 80

Leu Glu Gln Leu Asp Leu Ser Asp Asn Ala Gln Leu His Val Val Asp
85 90 95

Pro Thr Thr Phe His Gly Leu Gly His Leu His Thr Leu His Leu Asp
100 105 110

Arg Cys Gly Leu Arg Glu Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala
115 120 125

Ala Leu Gln Tyr Leu Tyr Leu Gln Asp Asn Asn Leu Gln Ala Leu Pro
130 135 140

Asp Asn Thr Phe Arg Asp Leu Gly Asn Leu Thr His Leu Phe Leu His
145 150 155 160

Gly Asn Arg Ile Pro Ser Val Pro Glu His Ala Phe Arg Gly Leu His
165 170 175

Ser Leu Asp Arg Leu Leu Leu His Gln Asn His Val Ala Arg Val His
180 185 190

Pro His Ala Phe Arg Asp Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe
195 200 205

Ala Asn Asn Leu Ser Met Leu Pro Ala Glu Val Leu Met Pro Leu Arg
210 215 220

Ser Leu Gln Tyr Leu Arg Leu Asn Asp Asn Pro Trp Val Cys Asp Cys
225 230 235 240

Arg Ala Arg Pro Leu Trp Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser
245 250 255

Ser Glu Val Pro Cys Asn Leu Pro Gln Arg Leu Ala Asp Arg Asp Leu
260 265 270

Lys Arg Leu Ala Ala Ser Asp Leu Glu Gly Cys Ala Val Ala Ser Gly
275 280 285

Pro Phe Arg Pro Ile Gln Thr Ser Gln Leu Thr Asp Glu Glu Leu Leu
290 295 300

Ser Leu Pro Lys Cys Cys Gln Pro Asp Ala Ala Asp Lys Ala Ser Val
305 310 315 320

Leu Glu Pro Gly Arg Pro Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg
325 330 335

Val Pro Pro Gly Asp Thr Pro Pro Gly Asn Gly Ser Gly Pro Arg His
340 345 350

Ile Asn Asp Ser Pro Phe Gly Thr Leu Pro Ser Ser Ala Glu Pro Pro
355 360 365

Leu Thr Ala Leu Arg Pro Gly Gly Ser Glu Pro Pro Gly Leu Pro Thr
370 375 380

Thr Gly Pro Arg Arg Arg Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg
385 390 395 400

Ser His Cys Arg Leu Gly Gln Ala Gly Ser Gly Ala Ser Gly Thr Gly
405 410 415

Asp Ala Glu Gly Ser Gly Ala Leu Pro Ala Leu Ala Cys Ser Leu Ala
420 425 430

Pro Leu Gly Leu Ala Leu Val Leu Trp Thr Val Leu Gly Pro Cys
435 440 445

<210> 98
<211> 18
<212> PRT
<213> Homo sapiens

<400> 98
Leu Pro Ala Leu Ala Cys Ser Leu Ala Pro Leu Gly Leu Ala Leu Val
1 5 10 15

Leu Trp

<210> 99
<211> 32
<212> PRT
<213> Homo sapiens

<400> 99
Pro Cys Pro Gly Ala Cys Val Cys Tyr Asn Glu Pro Lys Val Thr Thr
1   5   10   15

Ser Cys Pro Gln Gln Gly Leu Gln Ala Val Pro Thr Gly Ile Pro Ala
            20   25   30

<210> 100
<211> 24
<212> PRT
<213> Homo sapiens

<400> 100
Ser Ser Gln Arg Ile Phe Leu His Gly Asn Arg Ile Ser His Val Pro
1   5   10   15

Ala Ala Ser Phe Gln Ser Cys Arg
            20

<210> 101
<211> 24
<212> PRT
<213> Homo sapiens

<400> 101
Asn Leu Thr Ile Leu Trp Leu His Ser Asn Ala Leu Ala Arg Ile Asp
1   5   10   15

Ala Ala Ala Phe Thr Gly Leu Thr
            20

<210> 102
<211> 25
<212> PRT
<213> Homo sapiens

<400> 102
Leu Leu Glu Gln Leu Asp Leu Ser Asp Asn Ala Gln Leu His Val Val
1   5   10   15

Asp Pro Thr Thr Phe His Gly Leu Gly
            20   25

<210> 103
<211> 24
<212> PRT

<213> Homo sapiens

<400> 103
His Leu His Thr Leu His Leu Asp Arg Cys Gly Leu Arg Glu Leu Gly
1 5 10 15

Pro Gly Leu Phe Arg Gly Leu Ala
20

<210> 104
<211> 24
<212> PRT
<213> Homo sapiens

<400> 104
Ala Leu Gln Tyr Leu Tyr Leu Gln Asp Asn Asn Leu Gln Ala Leu Pro
1 5 10 15

Asp Asn Thr Phe Arg Asp Leu Gly
20

<210> 105
<211> 24
<212> PRT
<213> Homo sapiens

<400> 105
Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Pro Ser Val Pro
1 5 10 15

Glu His Ala Phe Arg Gly Leu His
20

<210> 106
<211> 24
<212> PRT
<213> Homo sapiens

<400> 106
Ser Leu Asp Arg Leu Leu Leu His Gln Asn His Val Ala Arg Val His
1 5 10 15

Pro His Ala Phe Arg Asp Leu Gly
20

<210> 107
<211> 24
<212> PRT
<213> Homo sapiens

<400> 107
Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Met Leu Pro
1 5 10 15

Ala Glu Val Leu Met Pro Leu Arg
20

<210> 108
<211> 51
<212> PRT
<213> Homo sapiens

<400> 108
Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp Ala Trp Leu
1 5 10 15

Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Asn Leu Pro Gln
20 25 30

Arg Leu Ala Asp Arg Asp Leu Lys Arg Leu Ala Ala Ser Asp Leu Glu
35 40 45

Gly Cys Ala
50

<210> 109
<211> 423
<212> PRT
<213> Homo sapiens

<400> 109
Cys Pro Gly Ala Cys Val Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser
1 5 10 15

Cys Pro Gln Gln Gly Leu Gln Ala Val Pro Thr Gly Ile Pro Ala Ser
20 25 30

Ser Gln Arg Ile Phe Leu His Gly Asn Arg Ile Ser His Val Pro Ala
35 40 45

Ala Ser Phe Gln Ser Cys Arg Asn Leu Thr Ile Leu Trp Leu His Ser
50 55 60

Asn Ala Leu Ala Arg Ile Asp Ala Ala Ala Phe Thr Gly Leu Thr Leu
65 70 75 80

Leu Glu Gln Leu Asp Leu Ser Asp Asn Ala Gln Leu His Val Val Asp
85 90 95

Pro Thr Thr Phe His Gly Leu Gly His Leu His Thr Leu His Leu Asp
100 105 110

Arg Cys Gly Leu Arg Glu Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala
115 120 125

Ala Leu Gln Tyr Leu Tyr Leu Gln Asp Asn Asn Leu Gln Ala Leu Pro
130 135 140

Asp Asn Thr Phe Arg Asp Leu Gly Asn Leu Thr His Leu Phe Leu His
145 150 155 160

Gly Asn Arg Ile Pro Ser Val Pro Glu His Ala Phe Arg Gly Leu His
165 170 175

Ser Leu Asp Arg Leu Leu Leu His Gln Asn His Val Ala Arg Val His
180 185 190

Pro His Ala Phe Arg Asp Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe
195 200 205

Ala Asn Asn Leu Ser Met Leu Pro Ala Glu Val Leu Met Pro Leu Arg
210 215 220

Ser Leu Gln Tyr Leu Arg Leu Asn Asp Asn Pro Trp Val Cys Asp Cys
225 230 235 240

Arg Ala Arg Pro Leu Trp Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser
245 250 255

Ser Glu Val Pro Cys Asn Leu Pro Gln Arg Leu Ala Asp Arg Asp Leu
260 265 270

Lys Arg Leu Ala Ala Ser Asp Leu Glu Gly Cys Ala Val Ala Ser Gly
275 280 285

Pro Phe Arg Pro Ile Gln Thr Ser Gln Leu Thr Asp Glu Glu Leu Leu
290 295 300

Ser Leu Pro Lys Cys Cys Gln Pro Asp Ala Ala Asp Lys Ala Ser Val
305 310 315 320

Leu Glu Pro Gly Arg Pro Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg
325 330 335

Val Pro Pro Gly Asp Thr Pro Pro Gly Asn Gly Ser Gly Pro Arg His
340 345 350

Ile Asn Asp Ser Pro Phe Gly Thr Leu Pro Ser Ser Ala Glu Pro Pro
355 360 365

Leu Thr Ala Leu Arg Pro Gly Gly Ser Glu Pro Pro Gly Leu Pro Thr
370 375 380

Thr Gly Pro Arg Arg Arg Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg
385 390 395 400

Ser His Cys Arg Leu Gly Gln Ala Gly Ser Gly Ala Ser Gly Thr Gly
405 410 415

Asp Ala Glu Gly Ser Gly Ala
420

<210> 110
<211> 1348
<212> DNA
<213> Homo sapiens

<400> 110

gccaaagaga catatccaag gttgagatta gtttccattt tctttgtact attttctgga 60
taataagaca ttagacattt gaagagatgg agaatgaaga tgggtatatg acgctgagtt 120
tcaagaatcg ttgtaaatcg aagcagaaat ctaaagattt ctccctatat ccacaatatt 180
attgtcttct gctcatattt ggatgcattg tgatccttat attcattatg acagggattg 240
acctgaagtt ctggcataaa aaatggattt tctcccagaa tgtaaacatc agcagtctat 300
caggacacaa ttacttgtgc ccaaatgact ggctgttgaa cgaagggaaa tgttactggt 360
tttcaacttc ttttaaaacg tggaaagaga gtcaacgtga ttgtacacag ctacaggcac 420
atttactggt gattcaaaat ttggatgagc tggagttcat acagaacagt ttaaaacctg 480
gacattttgg ttggattgga ctatatgtta cattccaagg gaacctatgg atgtggatag 540
atgaacactt tttagttcca gaattgtttt cagtgattgg accaactgat gacaggagct 600
gtgccgttat cacaggaaac tgggtgtatt ctgaagactg tagctccaca tttaagggca 660
tttgccagag agatgcgatc ttgacgcaca atggaaccag tggtgtgtaa atgtacaacc 720
aaatatagaa atactttgca tgttaaagca gagctagatt ttaaagactt aagattttta 780
gataaagttt ctaacagaaa gtttctgcta acagacatca tctaaatagg agaaaagtat 840
tttatcctga attgactata aagacaactt ctgaacagaa cttttactct atacttggat 900
ttctggtttg tcttttccat ggcattgaca agaaaagcta aataaaaaat tagtaattat 960
tttaatagtt atttaatagt ttgatttttt tgcatttaaa atagcataga ataaaacaac 1020
tttaaaggaa tgttatttag ctatatgtgc tatgtggtag attggaagga aagaagcagt 1080
atatgtacaa atataatatt tgaagcatgg aattctgaat ttttcatctg tgtattatag 1140
cctgaagtgt ttggtgggga gtgggtaatg agaaattacc tactgggtat aatgtacaat 1200
atttaggtga tggataaact aaaagctcag acttctccac tttgtgatat atccatgtaa 1260
caaaattatg cttgtaccct ttaaatgtat tcaaataaaa taaaataaag tcatgtggcc 1320
aaatattcaa aacaaaaaaa aaaaaaaa 1348

<210> 111
<211> 624
<212> DNA
<213> Homo sapiens

<400> 111
atggagaatg aagatgggta tatgacgctg agtttcaaga atcgttgtaa atcgaagcag 60
aaatctaaag atttctccct atatccacaa tattattgtc ttctgctcat atttggatgc 120
attgtgatcc ttatattcat tatgacaggg attgacctga agttctggca taaaaaaatg 180
gatttctccc agaatgtaaa catcagcagt ctatcaggac acaattactt gtgcccaaat 240
gactggctgt tgaacgaagg gaaatgttac tggttttcaa cttcttttaa aacgtggaaa 300
gagagtcaac gtgattgtac acagctacag gcacatttac tggtgattca aaatttggat 360
gagctggagt tcatacagaa cagtttaaaa cctggacatt ttggttggat tggactatat 420
gttacattcc aagggaacct atggatgtgg atagatgaac actttttagt tccagaattg 480
ttttcagtga ttggaccaac tgatgacagg agctgtgccg ttatcacagg aaactgggtg 540
tattctgaag actgtagctc cacatttaag ggcatttgcc agagagatgc gatcttgacg 600
cacaatggaa ccagtggtgt gtaa 624

<210> 112
<211> 207
<212> PRT
<213> Homo sapiens

<400> 112
Met Glu Asn Glu Asp Gly Tyr Met Thr Leu Ser Phe Lys Asn Arg Cys
1 5 10 15

Lys Ser Lys Gln Lys Ser Lys Asp Phe Ser Leu Tyr Pro Gln Tyr Tyr
20 25 30

Cys Leu Leu Leu Ile Phe Gly Cys Ile Val Ile Leu Ile Phe Ile Met
35 40 45

Thr Gly Ile Asp Leu Lys Phe Trp His Lys Lys Met Asp Phe Ser Gln

50 55 60

Asn Val Asn Ile Ser Ser Leu Ser Gly His Asn Tyr Leu Cys Pro Asn
65 70 75 80

Asp Trp Leu Leu Asn Glu Gly Lys Cys Tyr Trp Phe Ser Thr Ser Phe
85 90 95

Lys Thr Trp Lys Glu Ser Gln Arg Asp Cys Thr Gln Leu Gln Ala His
100 105 110

Leu Leu Val Ile Gln Asn Leu Asp Glu Leu Glu Phe Ile Gln Asn Ser
115 120 125

Leu Lys Pro Gly His Phe Gly Trp Ile Gly Leu Tyr Val Thr Phe Gln
130 135 140

Gly Asn Leu Trp Met Trp Ile Asp Glu His Phe Leu Val Pro Glu Leu
145 150 155 160

Phe Ser Val Ile Gly Pro Thr Asp Asp Arg Ser Cys Ala Val Ile Thr
165 170 175

Gly Asn Trp Val Tyr Ser Glu Asp Cys Ser Ser Thr Phe Lys Gly Ile
180 185 190

Cys Gln Arg Asp Ala Ile Leu Thr His Asn Gly Thr Ser Gly Val
195 200 205

<210> 113
<211> 157
<212> PRT
<213> Homo sapiens

<400> 113
Ile Asp Leu Lys Phe Trp His Lys Lys Met Asp Phe Ser Gln Asn Val
1 5 10 15

Asn Ile Ser Ser Leu Ser Gly His Asn Tyr Leu Cys Pro Asn Asp Trp
20 25 30

Leu Leu Asn Glu Gly Lys Cys Tyr Trp Phe Ser Thr Ser Phe Lys Thr
35 40 45

Trp Lys Glu Ser Gln Arg Asp Cys Thr Gln Leu Gln Ala His Leu Leu
50 55 60

Val Ile Gln Asn Leu Asp Glu Leu Glu Phe Ile Gln Asn Ser Leu Lys
65 70 75 80

Pro Gly His Phe Gly Trp Ile Gly Leu Tyr Val Thr Phe Gln Gly Asn
85 90 95

Leu Trp Met Trp Ile Asp Glu His Phe Leu Val Pro Glu Leu Phe Ser
100 105 110

Val Ile Gly Pro Thr Asp Asp Arg Ser Cys Ala Val Ile Thr Gly Asn

```
        115                 120                 125

Trp Val Tyr Ser Glu Asp Cys Ser Ser Thr Phe Lys Gly Ile Cys Gln
    130                 135                 140

Arg Asp Ala Ile Leu Thr His Asn Gly Thr Ser Gly Val
145                 150                 155


<210> 114
<211> 50
<212> PRT
<213> Homo sapiens

<400> 114
Met Glu Asn Glu Asp Gly Tyr Met Thr Leu Ser Phe Lys Asn Arg Cys
  1               5                  10                  15

Lys Ser Lys Gln Lys Ser Lys Asp Phe Ser Leu Tyr Pro Gln Tyr Tyr
             20                  25                  30

Cys Leu Leu Leu Ile Phe Gly Cys Ile Val Ile Leu Ile Phe Ile Met
         35                  40                  45

Thr Gly
     50


<210> 115
<211> 83
<212> PRT
<213> Homo sapiens

<400> 115
Ile Asp Leu Lys Phe Trp His Lys Lys Met Asp Phe Ser Gln Asn Val
  1               5                  10                  15

Asn Ile Ser Ser Leu Ser Gly His Asn Tyr Leu Cys Pro Asn Asp Trp
             20                  25                  30

Leu Leu Asn Glu Gly Lys Cys Tyr Trp Phe Ser Thr Ser Phe Lys Thr
         35                  40                  45

Trp Lys Glu Ser Gln Arg Asp Cys Thr Gln Leu Gln Ala His Leu Leu
     50                  55                  60

Val Ile Gln Asn Leu Asp Glu Leu Glu Phe Ile Gln Asn Ser Leu Lys
 65                  70                  75                  80

Pro Gly His


<210> 116
<211> 18
<212> PRT
<213> Homo sapiens
```

<400> 116
Phe Gly Trp Ile Gly Leu Tyr Val Thr Phe Gln Gly Asn Leu Trp Met
1   5   10   15

Trp Ile

<210> 117
<211> 56
<212> PRT
<213> Homo sapiens

<400> 117
Asp Glu His Phe Leu Val Pro Glu Leu Phe Ser Val Ile Gly Pro Thr
1   5   10   15

Asp Asp Arg Ser Cys Ala Val Ile Thr Gly Asn Trp Val Tyr Ser Glu
            20   25   30

Asp Cys Ser Ser Thr Phe Lys Gly Ile Cys Gln Arg Asp Ala Ile Leu
        35   40   45

Thr His Asn Gly Thr Ser Gly Val
    50   55

<210> 118
<211> 90
<212> PRT
<213> Homo sapiens

<400> 118
Arg Asp Cys Thr Gln Leu Gln Ala His Leu Leu Val Ile Gln Asn Leu
1   5   10   15

Asp Glu Leu Glu Phe Ile Gln Asn Ser Leu Lys Pro Gly His Phe Gly
            20   25   30

Trp Ile Gly Leu Tyr Val Thr Phe Gln Gly Asn Leu Trp Met Trp Ile
        35   40   45

Asp Glu His Phe Leu Val Pro Glu Leu Phe Ser Val Ile Gly Pro Thr
    50   55   60

Asp Asp Arg Ser Cys Ala Val Ile Thr Gly Asn Trp Val Tyr Ser Glu
65   70   75   80

Asp Cys Ser Ser Thr Phe Lys Gly Ile Cys
                85   90

<210> 119
<211> 22
<212> PRT
<213> Homo sapiens

<400> 119

Cys Ala Val Ile Thr Gly Asn Trp Val Tyr Ser Glu Asp Cys Ser Ser
1   5   10   15

Thr Phe Lys Gly Ile Cys
            20

<210> 120
<211> 2463
<212> DNA
<213> Homo sapiens

<400> 120
attttagtt tgttgaagtt cgtgactgct tcactctctc attcttagct tgaatttgga 60
aatgactttt gatgacctaa agatccagac tgtgaaggac cagcctgatg agaagtcaaa 120
tggaaaaaaa gctaaaggtc ttcagtttct ttactctcca tggtggtgcc tggctgctgc 180
gactctaggg gtcctttgcc tgggattagt agtgaccatt atggtgctgg gcatgcaatt 240
atcccaggtg tctgacctcc taacacaaga gcaagcaaac ctaactcacc agaaaaagaa 300
actggaggga cagatctcag cccggcaaca agcagaagaa gcttcacagg agtcagaaaa 360
cgaactcaag gaaatgatag aaacccttgc tcggaagctg aatgagaaat ccaaagagca 420
aatggaactt caccaccaga atctgaatct ccaagaaaca ctgaagagag tagcaaattg 480
ttcagctcct tgtccgcaag actggatctg gcatggagaa aactgttacc tattttcctc 540
gggctcattt aactgggaaa agagccaaga gaagtgcttg tctttggatg ccaagttgct 600
gaaaattaat agcacagctg atctggactt catccagcaa gcaatttcct attccagttt 660
tccattctgg atggggctgt ctcggaggaa ccccagctac ccatggctct gggaggacgg 720
ttctcctttg atgccccact tatttagagt ccgaggcgct gtctcccaga catacccttc 780
aggtacctgt gcatatatac aacgaggagc tgtttatgcg gaaaactgca ttttagctgc 840
cttcagtata tgtcagaaga aggcaaacct aagagcacag tgaatttgaa ggctctggaa 900
gaaaagaaaa aagtctttga gttttattct ggaatttaag ctattctttg tcacttgggt 960
gccaaacatg agagcccaga aaactgtcat ttagctggct gcagaactcc tttgcagaaa 1020
ctggggttcc aggtgcctgg cacctttatg tcaacatttt tgattctagc tatctgtatt 1080
atttcaccta gcttgtccca agcttccctg ccagcctgaa gtccattttc ccctttttat 1140
tttaaaattt gactcctctt caagcttgaa aaccctctga actcagtctt ctttacctca 1200
ttatcacctt cccctcacac tcctaaaatt gcatgaaaga cagaacatgg agaacttgct 1260
caagtgcagg cagagagcaa aaaggggaaa tatgtctggg aaaaagtgca cgtgaagaaa 1320
caaagaagga cagaggccat tccgaaatca agaaactcat gttcttaact ttaaaaaagg 1380
tatcaatcct tggtttttaa actgtggtcc atctccagac tctaccactt acggacagac 1440
agacagacag acacacacac acacacacac acacacattt tgggacaagt ggggagccca 1500
agaaagtaat tagtaagtga gtggtctttt ctgtaagcta atccacaacc tgttaccact 1560
tcctgaatca gttattattt cttcattttt ttttctacca gaggacagat taatagattt 1620
aacccttcac aacagttctt gttagaatca tgggatgtgt ggcccagagg taagaataga 1680
atttctttcc ctaaagaaca taccttttgt agatgaactc ttctcaactc tgttttgcta 1740
tgctataatt ccgaaacata caagacaaaa aaaatgaaga cactcaatct agaacaaact 1800
aagccaggta tgcaaatatc gctgaataga aacagatgga attagaaata tatcttctat 1860
ttttaggctt ctatttcctt tccacccact cttcacaggc tattctactc taaaggaagc 1920
ctttttattt tgctgcacac aatctagcag gaatcttttt tttttttta agagctgtgt 1980
catccttatg taggcaagag atgtttgctt ttgttaaaag ctttattgag atataattaa 2040
cataaaataa actgaacata tttaaagtgt actatttgat aagttttcac accttgtgga 2100
gaacatgcat actacaatta agagagtgaa catatccatc atccctcaaa gtgtcacaat 2160
gctcctcctg atgactcctc cccagaaaac caccaatcgg ctttcatttt gcattttgta 2220
gttttatgtg aatggaatca tatagtatgt cttttttttt tgtctggctt ctttcacttt 2280
gcataattat tttgagattc atatgtctcc atcttgatgc tcgtatgaat tcattctttt 2340
aaatgttgaa tattcccttg tatggatata ccacaattca tttacccatt tacttgttga 2400
tgacatttgg gttgttttag ttttgggata ttacaaataa agctgctgtg aacatttgtg 2460
tac 2463

<210> 121
<211> 822

<212> DNA
<213> Homo sapiens

<400> 121
atgacttttg atgacctaaa gatccagact gtgaaggacc agcctgatga gaagtcaaat 60
ggaaaaaaag ctaaaggtct tcagtttctt tactctccat ggtggtgcct ggctgctgcg 120
actctagggg tcctttgcct gggattagta gtgaccatta tggtgctggg catgcaatta 180
tcccaggtgt ctgacctcct aacacaagag caagcaaacc taactcacca gaaaaagaaa 240
ctggagggac agatctcagc ccggcaacaa gcagaagaag cttcacagga gtcagaaaac 300
gaactcaagg aaatgataga aacccttgct cggaagctga atgagaaatc caaagagcaa 360
atggaacttc accaccagaa tctgaatctc caagaaacac tgaagagagt agcaaattgt 420
tcagctcctt gtccgcaaga ctggatctgg catggagaaa actgttacct attttcctcg 480
ggctcattta actgggaaaa gagccaagag aagtgcttgt ctttggatgc caagttgctg 540
aaaattaata gcacagctga tctggacttc atccagcaag caatttccta ttccagtttt 600
ccattctgga tggggctgtc tcggaggaac cccagctacc catggctctg ggaggacggt 660
tctccctttga tgccccactt atttagagtc cgaggcgctg tctcccagac ataccttca 720
ggtacctgtg catatataca acgaggagct gtttatgcgg aaaactgcat tttagctgcc 780
ttcagtatat gtcagaagaa ggcaaaccta agagcacagt ga 822

<210> 122
<211> 273
<212> PRT
<213> Homo sapiens

<400> 122
Met Thr Phe Asp Asp Leu Lys Ile Gln Thr Val Lys Asp Gln Pro Asp
1 5 10 15

Glu Lys Ser Asn Gly Lys Lys Ala Lys Gly Leu Gln Phe Leu Tyr Ser
20 25 30

Pro Trp Trp Cys Leu Ala Ala Ala Thr Leu Gly Val Leu Cys Leu Gly
35 40 45

Leu Val Val Thr Ile Met Val Leu Gly Met Gln Leu Ser Gln Val Ser
50 55 60

Asp Leu Leu Thr Gln Glu Gln Ala Asn Leu Thr His Gln Lys Lys Lys
65 70 75 80

Leu Glu Gly Gln Ile Ser Ala Arg Gln Gln Ala Glu Glu Ala Ser Gln
85 90 95

Glu Ser Glu Asn Glu Leu Lys Glu Met Ile Glu Thr Leu Ala Arg Lys
100 105 110

Leu Asn Glu Lys Ser Lys Glu Gln Met Glu Leu His His Gln Asn Leu
115 120 125

Asn Leu Gln Glu Thr Leu Lys Arg Val Ala Asn Cys Ser Ala Pro Cys
130 135 140

Pro Gln Asp Trp Ile Trp His Gly Glu Asn Cys Tyr Leu Phe Ser Ser
145 150 155 160

Gly Ser Phe Asn Trp Glu Lys Ser Gln Glu Lys Cys Leu Ser Leu Asp
165 170 175

Ala Lys Leu Leu Lys Ile Asn Ser Thr Ala Asp Leu Asp Phe Ile Gln
180 185 190

Gln Ala Ile Ser Tyr Ser Ser Phe Pro Phe Trp Met Gly Leu Ser Arg
195 200 205

Arg Asn Pro Ser Tyr Pro Trp Leu Trp Glu Asp Gly Ser Pro Leu Met
210 215 220

Pro His Leu Phe Arg Val Arg Gly Ala Val Ser Gln Thr Tyr Pro Ser
225 230 235 240

Gly Thr Cys Ala Tyr Ile Gln Arg Gly Ala Val Tyr Ala Glu Asn Cys
245 250 255

Ile Leu Ala Ala Phe Ser Ile Cys Gln Lys Lys Ala Asn Leu Arg Ala
260 265 270

Gln

<210> 123
<211> 1196
<212> DNA
<213> Homo sapiens

<400> 123
agcatctcta gacctagagg ttttctctat ttctcctttt cactgtgacc caggaaataa 60
ttttcagaag taaaaaaatc tcatctgaga ctctgcaaca ggcaccagag agtgaggaag 120
aactttgagt aacagaaact gctccaattt ctcatccgca tctcacatct ctgtgtcaac 180
tatccttcta tcccatttat tctggtatta gatatgttgt cagtgtctct tgttaggtag 240
agaaatcagc agtcagatct taagaccatt tggtaggtgc atcaggaatt gacacgcagg 300
ccagtttcc agtcctacat gtattctttt ctctgtatcc tgcctctctt gctcttggct 360
tcctgccttc tctcctactc atttttagaa cagtctagat gcaggcagct agaagagtta 420
tttcctccaa gctgtctagg aaaagggaca attaaagaga gattctgcac ttattatgat 480
ataaaaaaag aaaaacaatg agagaacata gtactctttt acctgtgacg ttaaatggga 540
gtcacacagg tttagctatc tgttctagga gtggatgaac agactattct cccatgtcac 600
ttctttcctc ctggacacct tcagggagac agctgggtga agaatcatac ttctgacctc 660
tgtcaaacag ggtcagatgc cgcagaggtt ctgagatgat aaaggaagtg acagagggaa 720
cctgaggtac cacatttctg atttgtcatg aaagtcttac cttgcttaag atgacttttt 780
taatgttcct ttcagggaaa atgccaagtg gaataaaaac caacatcaag tcggcttcca 840
tgcatcccta ccagcggtga gtgtggctgg caacctcgac tccctggtgc tctttgcaga 900
gttgggcagt gaaattacct tttgctcaag gctcacctag atgggtacaa taaaaagaac 960
atgggctttc agcagcagac aaatcccact tccaccactg actagctgtg tgaccttgga 1020
caagtgacct aatttttctg agcctgtttc tcatttgtaa atggtgata[illegible] tacctacctc 1080
atagggttgt tgtgaggatt aaaatgagga aatgaatgta aagcacttag tacagtatat 1140
gaaataatgg gtattcaata aatgatagtt tctacagaaa aaaaaaaaaa aaaaaa 1196

<210> 124
<211> 183
<212> DNA
<213> Homo sapiens

<400> 124
atgtattctt ttctctgtat cctgcctctc ttgctcttgg cttcctgcct tctctcctac 60
tcatttttag aacagtctag atgcaggcag ctagaagagt tatttcctcc aagctgtcta 120
ggaaaaggga caattaaaga gagattctgc acttattatg atataaaaaa agaaaaacaa 180

tga 183

<210> 125
<211> 60
<212> PRT
<213> Homo sapiens

<400> 125
Met Tyr Ser Phe Leu Cys Ile Leu Pro Leu Leu Leu Leu Ala Ser Cys
1 5 10 15

Leu Leu Ser Tyr Ser Phe Leu Glu Gln Ser Arg Cys Arg Gln Leu Glu
20 25 30

Glu Leu Phe Pro Pro Ser Cys Leu Gly Lys Gly Thr Ile Lys Glu Arg
35 40 45

Phe Cys Thr Tyr Tyr Asp Ile Lys Lys Glu Lys Gln
50 55 60

<210> 126
<211> 19
<212> PRT
<213> Homo sapiens

<400> 126
Met Tyr Ser Phe Leu Cys Ile Leu Pro Leu Leu Leu Leu Ala Ser Cys
1 5 10 15

Leu Leu Ser

<210> 127
<211> 41
<212> PRT
<213> Homo sapiens

<400> 127
Tyr Ser Phe Leu Glu Gln Ser Arg Cys Arg Gln Leu Glu Glu Leu Phe
1 5 10 15

Pro Pro Ser Cys Leu Gly Lys Gly Thr Ile Lys Glu Arg Phe Cys Thr
20 25 30

Tyr Tyr Asp Ile Lys Lys Glu Lys Gln
35 40

<210> 128
<211> 3649
<212> DNA
<213> Homo sapiens

<400> 128
ggcgggctga ggagccgccg ggctcgggct ccgcgcgggg atgtgtctgg ccaccgcgta 60
ctttcgaggg cgcctcaaca cccggggact gatgcccttg gccagacgca cccccaactg 120

```
ctgtctgctc aggtgtagcc cgcgtgactg agtcaccgct gctccagctg tttcacgtcg 180
ccgttccctt tactgaatag tttgatgggg cgccgggcgg atgacagcgg gaacggttgt 240
gatcactggc ggaatcctag ctacggtgat cctcctctgc atcattgccg tcctgtgcta 300
ctgcaggctc cagtattact gctgcaagaa gagcggaacc gaggttgcag acgaggagga 360
ggagcgggag cacgaccttc ccacgcatcc cagaggcccc acctgcaatg cctgcagctc 420
ccaagccctg gacggcagag gcagcctggc gcctctcacc agcgagccct gcagccagcc 480
ctgtggggtg gccgcgagcc actgcactac ctgctcccca tacagctccc cctttacat 540
acggacggct gacatggtgc ccaatggggg tggaggcgag aggctctcct ttgctcccac 600
atactacaaa gaggggggac cccatccct caaattggca gcaccccaga gttacccggt 660
gacctggcca ggctctgggc gtgaggcctt caccaatcca agggctatta gtacagacgt 720
gtaaatcctt ccaccccgac ccgcacacac acccacactg ctgccctggc gggggccatg 780
ggggtgatga atgaccctcc aacagcccca catgggttgt ttctgtttct ttggcttttc 840
tcgctccgca gtggagggtt tactaggatt taagcttttg agtgcattga gaaccaagac 900
agggcctggc tccaactctg tgggccagag gtgggggact gctaggtcga gtctgcagct 960
tcgccagttt cttggttggg acactcctct ggcagcccca gcaccaccac aacccccttgc 1020
agtgtgcccc agtcccctgg attcgctgga ctgcaaaaag gagcaccagg gaaagctcag 1080
caaaggctca ggaggcctcc cgggtcctcg ggagatgaag catccgtgcc taggaaaagg 1140
ggacagaggg caaggagaga agtgagagct acaattccag cattggagaa gcgaggggcg 1200
gggcgtcaac ggcactgctt caggacgcgc ttgctgaaac gactccaaca gctagttcac 1260
agcccagctt tgtacgttgg ttaccatagc tactgctgtc actgtagctg ctcccgtagg 1320
gacgttgatt cctgaacgta tcacatctca cctgcccct tcctcgtggg acgtgtcaag 1380
ttgactttaa agcctaaggt ggcttgtggg gactgcacca gaaagtgtct aaccttgtgt 1440
tccccgaaat cctttctctg acttaggaac cagcgccccc tgctggagaa gttttttcat 1500
tttgttagcc gacttctggt agcttagcaa agagaccagc tgactgtctc tcggcccagc 1560
ccaccacgcc gagtagcttg tgtggatgca gtcctgtgag ggtgtgcata accgttccca 1620
ggtgtacgca cgcgcgcgcg cgcacacaca cacacacaca cacacacaca cacacagtaa 1680
ccacgggtag gcaccagggg ctttgtacat tcaggagctg cagagaggaa aggtctccca 1740
tgtaccagag aaggaaaagt ctgatcccga aacagcttga acgaaaggtg gttaacacgt 1800
gaccaatccc catggcagac agcgggctgg tgaacaggaa gagcacagaa tctctgtggg 1860
gcgatgctgc agcggctgaa accacgtcaa gtcccccaga gcccggcatt ctatgtaagc 1920
atccacctgt gacgccgtgg agcaggatct tcattcatct caccagagca acagcccagt 1980
cctgaggatg ggagggacgt gacttagcct tgggattcca cgctcgccgc cctccccttgg 2040
ccgcccctact ccccagtttc ctggcgatga gccatttatt gtgtctcaca ccagcctcat 2100
ttccaactcc gacttcctac ttagaatgtg gcagtggctt tgtgagtgag aatcctgatg 2160
agttttccag tggcctctgc agaagctggc ccccttctaa gggcatagct aacccccttc 2220
ccactcctga gtcactgaat tcgaggaggg tggggcaggg aaagggctgg cccccctgct 2280
gaaatcttga ttctgcattt gaggaaggtc agggtgccgc tgggggagca gggagcgggg 2340
agggaggaag aggggaatgc acatcttttg cgccactatt aaggcacctg tcaccttaca 2400
tctcaatctg gaatcaaagt gccctgggtt gagaagcaga cctgggctct ggtcacactt 2460
tggtcactca ttagctctag gaccagtcac taatctctga gactccattt tctccaggga 2520
aacgaggctt gcccagatag actaacaacc ctttagtgct ccaagaagtg agaatttccg 2580
aaaaagatct gccggcccag agcactccct cttgccctaa tcctggcagg gtctggatgt 2640
atttactcaa ctctgatctt gcgtgcaggg atgcttgctg atggatgaaa gaccacaata 2700
aaacaagatt agcagcaaaa caattttaat tgtcagtgaa aagcaaacat gtttgtcaga 2760
ctctcctgtg gagtcatttc aaatggtaac tcttctcttg taccaggaag cagctgtagg 2820
tggggtaatg ggatctgtct cctggtgcag aggggcagct tatggcggca gcagcttcaa 2880
caagcccgtc ccacggccag ccctaactgg gccaagtgca agctcctgat gagctgctgg 2940
gcctggcagc gttccccact tcctagtgcc ttgctcccct ggaattccac agccaatcaa 3000
ccacagcagc cgtgtcctcc ttgccagaaa taacatcatt cgtacgtcct gcctttgtaa 3060
aaatcaagac ccagttaacc atcacccact tcttgcccct gattgatggg gtccctgagt 3120
acaagcatcc tgtgcaatcc cctccctcct accgctttcc tagctaagcc ctgctagccc 3180
cacaaatcag aatctgggtt tggggtgctg gagagatggt gggccccagc tcttccccac 3240
atagcccacg ctgctcaggg cacggtctcc aaccaggctc aagattgcct gcctgcccgc 3300
tgcagccgca gggccccacc ctgtggaaaa ggaagcttcc cccactcaac tgcactgtcc 3360
ttttctgctg tctagtctgt gctaagggac actatttgta cacttttcaa acggtgccct 3420
aaattggaag agaaggaggg gccaggaaca cagccctctc cgccgtcttt ctccacctcc 3480
acttcaccac tcctgctcta gacttcttgt cttgcatctt tgataacttg gagtcataac 3540
cgagtggatg tcggaataaa tgagagattt gaagtngaaa aaaaaaaaaa aaaaaaaaaa 3600
```

aaaaaaaaaa aaaaaaaaaa aaaaaaaata aaaaaaaaaa aaaaaaaaa 3649

<210> 129
<211> 504
<212> DNA
<213> Homo sapiens

<400> 129
atgacagcgg gaacggttgt gatcactggc ggaatcctag ctacggtgat cctcctctgc 60
atcattgccg tcctgtgcta ctgcaggctc cagtattact gctgcaagaa gagcggaacc 120
gaggttgcag acgaggagga ggagcgggag cacgaccttc ccacgcatcc cagaggcccc 180
acctgcaatg cctgcagctc ccaagccctg gacggcagag gcagcctggc gcctctcacc 240
agcgagccct gcagccagcc ctgtggggtg gccgcgagcc actgcactac ctgctcccca 300
tacagctccc ccttttacat acggacggct gacatggtgc ccaatggggg tggaggcgag 360
aggctctcct ttgctcccac atactacaaa gagggggac ccccatccct caaattggca 420
gcaccccaga gttacccggt gacctggcca ggctctgggc gtgaggcctt caccaatcca 480
agggctatta gtacagacgt gtaa 504

<210> 130
<211> 167
<212> PRT
<213> Homo sapiens

<400> 130
Met Thr Ala Gly Thr Val Val Ile Thr Gly Gly Ile Leu Ala Thr Val
1   5   10   15

Ile Leu Leu Cys Ile Ile Ala Val Leu Cys Tyr Cys Arg Leu Gln Tyr
            20   25   30

Tyr Cys Cys Lys Lys Ser Gly Thr Glu Val Ala Asp Glu Glu Glu Glu
        35   40   45

Arg Glu His Asp Leu Pro Thr His Pro Arg Gly Pro Thr Cys Asn Ala
    50   55   60

Cys Ser Ser Gln Ala Leu Asp Gly Arg Gly Ser Leu Ala Pro Leu Thr
65   70   75   80

Ser Glu Pro Cys Ser Gln Pro Cys Gly Val Ala Ala Ser His Cys Thr
                85   90   95

Thr Cys Ser Pro Tyr Ser Ser Pro Phe Tyr Ile Arg Thr Ala Asp Met
            100   105   110

Val Pro Asn Gly Gly Gly Gly Glu Arg Leu Ser Phe Ala Pro Thr Tyr
        115   120   125

Tyr Lys Glu Gly Gly Pro Pro Ser Leu Lys Leu Ala Ala Pro Gln Ser
    130   135   140

Tyr Pro Val Thr Trp Pro Gly Ser Gly Arg Glu Ala Phe Thr Asn Pro
145   150   155   160

Arg Ala Ile Ser Thr Asp Val
                165

<210> 131
<211> 26
<212> PRT
<213> Homo sapiens

<400> 131
Met Thr Ala Gly Thr Val Val Ile Thr Gly Gly Ile Leu Ala Thr Val
1   5   10   15

Ile Leu Leu Cys Ile Ile Ala Val Leu Cys
            20   25

<210> 132
<211> 141
<212> PRT
<213> Homo sapiens

<400> 132
Tyr Cys Arg Leu Gln Tyr Tyr Cys Cys Lys Lys Ser Gly Thr Glu Val
1   5   10   15

Ala Asp Glu Glu Glu Glu Arg Glu His Asp Leu Pro Thr His Pro Arg
            20   25   30

Gly Pro Thr Cys Asn Ala Cys Ser Ser Gln Ala Leu Asp Gly Arg Gly
        35   40   45

Ser Leu Ala Pro Leu Thr Ser Glu Pro Cys Ser Gln Pro Cys Gly Val
    50   55   60

Ala Ala Ser His Cys Thr Thr Cys Ser Pro Tyr Ser Ser Pro Phe Tyr
65   70   75   80

Ile Arg Thr Ala Asp Met Val Pro Asn Gly Gly Gly Gly Glu Arg Leu
                85   90   95

Ser Phe Ala Pro Thr Tyr Tyr Lys Glu Gly Gly Pro Pro Ser Leu Lys
            100   105   110

Leu Ala Ala Pro Gln Ser Tyr Pro Val Thr Trp Pro Gly Ser Gly Arg
        115   120   125

Glu Ala Phe Thr Asn Pro Arg Ala Ile Ser Thr Asp Val
    130   135   140

<210> 133
<211> 7
<212> PRT
<213> Homo sapiens

<400> 133
Arg Arg His His Pro Leu Gln
1   5

<210> 134

<211> 6
<212> PRT
<213> Homo sapiens

<400> 134
Thr Val Leu Gly Pro Cys
1 5

<210> 135
<211> 6
<212> PRT
<213> Homo sapiens

<400> 135
Thr Val Leu Gly Pro Cys
1 5

<210> 136
<211> 813
<212> DNA
<213> Homo sapiens

<400> 136
atgcactatt atagatactc taacgccagg gtcagctgct ggtacaagta cctccttttc 60
agctacaaca tcatcttctg gttggctgga gttgtcttcc ttggagtcgg gctgtgggca 120
tggagcgaaa agggtgtgct gtccgacctc accaaagtga cccggatgca tggaatcgac 180
cctgtggtgc tggtcctgat ggtgggcgtg gtgatgttca ccctggggtt cgccggctgc 240
gtgggggctc tgcgggagaa tatctgcttg ctcaactttt tctgtggcac catcgtgctc 300
atcttcttcc tggagctggc tgtggccgtg ctggccttcc tgttccagga ctgggtgagg 360
gaccggttcc gggagttctt cgagagcaac atcaagtcct accgggacga tatcgatctg 420
caaaacctca tcgactccct tcagaaagct aaccagtgct gtggcgcata tggccctgaa 480
gactgggacc tcaacgtcta cttcaattgc agcggtgcca gctacagccg agagaagtgc 540
ggggtccccct tctcctgctg cgtgccagat cctgcgcaaa aagttgtgaa cacacagtgt 600
ggatatgatg tcaggattca gctgaagagc aagtgggatg agtccatctt cacgaaaggc 660
tgcatccagg cgctggaaag ctggctcccg cggaacattt acattgtggc tggcgtcttc 720
atcgccatct cgctgttgca gatatttggc atcttcctgg caaggacgct gatctcagac 780
atcgaggcag tgaaggccgg ccatcacttc tga 813

<210> 137
<211> 270
<212> PRT
<213> Homo sapiens

<400> 137
Met His Tyr Tyr Arg Tyr Ser Asn Ala Arg Val Ser Cys Trp Tyr Lys
1 5 10 15

Tyr Leu Leu Phe Ser Tyr Asn Ile Ile Phe Trp Leu Ala Gly Val Val
20 25 30

Phe Leu Gly Val Gly Leu Trp Ala Trp Ser Glu Lys Gly Val Leu Ser
35 40 45

Asp Leu Thr Lys Val Thr Arg Met His Gly Ile Asp Pro Val Val Leu
50 55 60

Val Leu Met Val Gly Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys
65 70 75 80

Val Gly Ala Leu Arg Glu Asn Ile Cys Leu Leu Asn Phe Phe Cys Gly
85 90 95

Thr Ile Val Leu Ile Phe Phe Leu Glu Leu Ala Val Ala Val Leu Ala
100 105 110

Phe Leu Phe Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu
115 120 125

Ser Asn Ile Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile
130 135 140

Asp Ser Leu Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu
145 150 155 160

Asp Trp Asp Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser
165 170 175

Arg Glu Lys Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala
180 185 190

Gln Lys Val Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu
195 200 205

Lys Ser Lys Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala
210 215 220

Leu Glu Ser Trp Leu Pro Arg Asn Ile Tyr Ile Val Ala Gly Val Phe
225 230 235 240

Ile Ala Ile Ser Leu Leu Gln Ile Phe Gly Ile Phe Leu Ala Arg Thr
245 250 255

Leu Ile Ser Asp Ile Glu Ala Val Lys Ala Gly His His Phe
260 265 270

<210> 138
<211> 813
<212> DNA
<213> Homo sapiens

<400> 138
atgcactatt atagatactc taacgccaag gtcagctgct ggtacaagta cctcctttac 60
agctacaaca tcatcttctg gttggctgga gttgtcttcc ttggagtcgg gctgtgggca 120
tggagcgaaa agggtgtgct gtccgacctc accaaagtga cccggatgca tggaatcgac 180
cctgtggtgc tggtcctgat ggtgggcgtg gtgatgttca ccctggggtt cgccggctgc 240
gtgggggctc tgcgggagaa tatctgcttg ctcaactttt tctgtggcac catcgtgctc 300
atcttcttcc tggagctggc tgtggccgtg ctggccttcc tgttccagga ctgggtgagg 360
gaccggttcc gggagttctt cgagagcaac atcaagtcct accgggacga tatcgatctg 420
caaaacctca tcgactccct tcagaaagct aaccagtgct gtggcgcata tggccctgaa 480
gactgggacc tcaacgtcta cttcaattgc agcggtgcca gctacagccg agagaagtgc 540
ggggtcccct tctcctgctg cgtgccagat cctgcgcaaa aagttgtgaa cacacagtgt 600
ggatatgatg tcaggattca gctgaagagc aagtgggatg agtccatctt cacgaaaggc 660
tgcatccagg cgctggaaag ctggctcccg cggaacattt acattgtggc tggcgtcttc 720

atcgccatct cgctgttgca gatatttggc atcttcctgg caaggacgct gatctcagac 780
atcgaggcag tgaaggccgg ccatcacttc tga 813

<210> 139
<211> 270
<212> PRT
<213> Homo sapiens

<400> 139

Met His Tyr Tyr Arg Tyr Ser Asn Ala Lys Val Ser Cys Trp Tyr Lys
1 5 10 15

Tyr Leu Leu Tyr Ser Tyr Asn Ile Ile Phe Trp Leu Ala Gly Val Val
20 25 30

Phe Leu Gly Val Gly Leu Trp Ala Trp Ser Glu Lys Gly Val Leu Ser
35 40 45

Asp Leu Thr Lys Val Thr Arg Met His Gly Ile Asp Pro Val Val Leu
50 55 60

Val Leu Met Val Gly Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys
65 70 75 80

Val Gly Ala Leu Arg Glu Asn Ile Cys Leu Leu Asn Phe Phe Cys Gly
85 90 95

Thr Ile Val Leu Ile Phe Phe Leu Glu Leu Ala Val Ala Val Leu Ala
100 105 110

Phe Leu Phe Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu
115 120 125

Ser Asn Ile Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile
130 135 140

Asp Ser Leu Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu
145 150 155 160

Asp Trp Asp Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser
165 170 175

Arg Glu Lys Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala
180 185 190

Gln Lys Val Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu
195 200 205

Lys Ser Lys Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala
210 215 220

Leu Glu Ser Trp Leu Pro Arg Asn Ile Tyr Ile Val Ala Gly Val Phe
225 230 235 240

Ile Ala Ile Ser Leu Leu Gln Ile Phe Gly Ile Phe Leu Ala Arg Thr
245 250 255

Leu Ile Ser Asp Ile Glu Ala Val Lys Ala Gly His His Phe

260 265 270

<210> 140
<211> 813
<212> DNA
<213> Homo sapiens

<400> 140
atgcactatt atagatactc taacgccaag gtcagctgct ggtacaagta cctccttttc 60
agctacaaca tcatcttctg gttggctgga gttgtcttcc ttggagtcgg gctgtgggta 120
tggagcgaaa agggtgtgct gtccgacctc accaaagtga cccggatgca tggaatcgac 180
cctgtggtgc tggtcctgat ggtgggcgtg gtgatgttca ccctggggtt cgccggctgc 240
gtgggggctc tgcgggagaa tatctgcttg ctcaactttt tctgtggcac catcgtgctc 300
atcttcttcc tggagctggc tgtggccgtg ctggccttcc tgttccagga ctgggtgagg 360
gaccggttcc gggagttctt cgagagcaac atcaagtcct accgggacga tatcgatctg 420
caaaacctca tcgactccct tcagaaagct aaccagtgct gtggcgcata tggccctgaa 480
gactgggacc tcaacgtcta cttcaattgc agcggtgcca gctacagccg agagaagtgc 540
ggggtcccct tctcctgctg cgtgccagat cctgcgcaaa aagttgtgaa cacacagtgt 600
ggatatgatg tcaggattca gctgaagagc aagtgggatg agtccatctt cacgaaaggc 660
tgcatccagg cgctggaaag ctggctcccg cggaacattt acattgtggc tggcgtcttc 720
atcgccatct cgctgttgca gatatttggc atcttcctgg caaggacgct gatctcagac 780
atcgaggcag tgaaggccgg ccatcacttc tga 813

<210> 141
<211> 270
<212> PRT
<213> Homo sapiens

<400> 141
Met His Tyr Tyr Arg Tyr Ser Asn Ala Lys Val Ser Cys Trp Tyr Lys
1 5 10 15

Tyr Leu Leu Phe Ser Tyr Asn Ile Ile Phe Trp Leu Ala Gly Val Val
20 25 30

Phe Leu Gly Val Gly Leu Trp Val Trp Ser Glu Lys Gly Val Leu Ser
35 40 45

Asp Leu Thr Lys Val Thr Arg Met His Gly Ile Asp Pro Val Val Leu
50 55 60

Val Leu Met Val Gly Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys
65 70 75 80

Val Gly Ala Leu Arg Glu Asn Ile Cys Leu Leu Asn Phe Phe Cys Gly
85 90 95

Thr Ile Val Leu Ile Phe Phe Leu Glu Leu Ala Val Ala Val Leu Ala
100 105 110

Phe Leu Phe Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu
115 120 125

Ser Asn Ile Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile
130 135 140

Asp Ser Leu Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu

145 150 155 160

Asp Trp Asp Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser
165 170 175

Arg Glu Lys Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala
180 185 190

Gln Lys Val Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu
195 200 205

Lys Ser Lys Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala
210 215 220

Leu Glu Ser Trp Leu Pro Arg Asn Ile Tyr Ile Val Ala Gly Val Phe
225 230 235 240

Ile Ala Ile Ser Leu Leu Gln Ile Phe Gly Ile Phe Leu Ala Arg Thr
245 250 255

Leu Ile Ser Asp Ile Glu Ala Val Lys Ala Gly His His Phe
260 265 270

<210> 142
<211> 813
<212> DNA
<213> Homo sapiens

<400> 142
atgcactatt atagatactc taacgccaag gtcagctgct ggtacaagta cctccttttc 60
agctacaaca tcatcttctg gttggctgga gttgtcttcc ttggagtcgg gctgtgggca 120
tggagcgaaa agggtgtgct gtccgacctc accaaagtga cccggatgca tggaatcgag 180
cctgtggtgc tggtcctgat ggtgggcgtg gtgatgttca ccctggggtt cgccggctgc 240
gtgggggctc tgcgggagaa tatctgcttg ctcaactttt tctgtggcac catcgtgctc 300
atcttcttcc tggagctggc tgtggccgtg ctggccttcc tgttccagga ctgggtgagg 360
gaccggttcc gggagttctt cgagagcaac atcaagtcct accgggacga tatcgatctg 420
caaaacctca tcgactccct tcagaaagct aaccagtgct gtggcgcata tggccctgaa 480
gactgggacc tcaacgtcta cttcaattgc agcggtgcca gctacagccg agagaagtgc 540
ggggtcccct tctcctgctg cgtgccagat cctgcgcaaa aagttgtgaa cacacagtgt 600
ggatatgatg tcaggattca gctgaagagc aagtgggatg agtccatctt cacgaaaggc 660
tgcatccagg cgctggaaag ctggctcccg cggaacattt acattgtggc tggcgtcttc 720
atcgccatct cgctgttgca gatatttggc atcttcctgg caaggacgct gatctcagac 780
atcgaggcag tgaaggccgg ccatcacttc tga 813

<210> 143
<211> 270
<212> PRT
<213> Homo sapiens

<400> 143
Met His Tyr Tyr Arg Tyr Ser Asn Ala Lys Val Ser Cys Trp Tyr Lys
1 5 10 15

Tyr Leu Leu Phe Ser Tyr Asn Ile Ile Phe Trp Leu Ala Gly Val Val
20 25 30

Phe Leu Gly Val Gly Leu Trp Ala Trp Ser Glu Lys Gly Val Leu Ser

35 40 45

Asp Leu Thr Lys Val Thr Arg Met His Gly Ile Glu Pro Val Val Leu
50 55 60

Val Leu Met Val Gly Val Val Met Phe Thr Leu Gly Phe Ala Gly Cys
65 70 75 80

Val Gly Ala Leu Arg Glu Asn Ile Cys Leu Leu Asn Phe Phe Cys Gly
85 90 95

Thr Ile Val Leu Ile Phe Phe Leu Glu Leu Ala Val Ala Val Leu Ala
100 105 110

Phe Leu Phe Gln Asp Trp Val Arg Asp Arg Phe Arg Glu Phe Phe Glu
115 120 125

Ser Asn Ile Lys Ser Tyr Arg Asp Asp Ile Asp Leu Gln Asn Leu Ile
130 135 140

Asp Ser Leu Gln Lys Ala Asn Gln Cys Cys Gly Ala Tyr Gly Pro Glu
145 150 155 160

Asp Trp Asp Leu Asn Val Tyr Phe Asn Cys Ser Gly Ala Ser Tyr Ser
165 170 175

Arg Glu Lys Cys Gly Val Pro Phe Ser Cys Cys Val Pro Asp Pro Ala
180 185 190

Gln Lys Val Val Asn Thr Gln Cys Gly Tyr Asp Val Arg Ile Gln Leu
195 200 205

Lys Ser Lys Trp Asp Glu Ser Ile Phe Thr Lys Gly Cys Ile Gln Ala
210 215 220

Leu Glu Ser Trp Leu Pro Arg Asn Ile Tyr Ile Val Ala Gly Val Phe
225 230 235 240

Ile Ala Ile Ser Leu Leu Gln Ile Phe Gly Ile Phe Leu Ala Arg Thr
245 250 255

Leu Ile Ser Asp Ile Glu Ala Val Lys Ala Gly His His Phe
260 265 270

<210> 144
<211> 693
<212> DNA
<213> Homo sapiens

<400> 144
atgccctgga ccatcttgct ctttgcagct ggctccttgg cgatcccagc accatccatc 60
cggctggcgc ccccgtaccc aagcagccaa gaggacccca tccacatcgc atgcatggcc 120
cctgggaact tcccgggggc gaatttcaca ctgtatcgag gggggcaggt ggtccagctc 180
ctgcaggccc ccacggacca gcgcggggtg acatttaacc tgagcggcgg cagcagcaag 240
gctccagggg gacccttcca ctgccagtat ggagtgttag gtgagctcaa ccagtcccag 300
ctgtcagacc tcagcgagcc cgtgaacgtc tccttcccag tgcccacttg gatcttggtg 360
ctctccctga gcctggctgg tgccctcttc ctccttgctg ggctggtggc tgttgccctg 420

```
gtggtcagaa aagttaaact cagaaattta cagaagaaaa gagatcgaga atcctgctgg 480
gcccagatta acttcgacag cacagacatg tccttcgata actccctgtt taccgtctcc 540
gcgaaaacga tgccagaaga agacccggcc accttggatg atcactcagg caccactgcc 600
acccccagca actccaggac ccggaagagg cccacttcca cgtcctcctc gcctgagacc 660
cccgaattca gcactttccg ggcctgccag tga                              693

<210> 145
<211> 230
<212> PRT
<213> Homo sapiens

<400> 145
Met Pro Trp Thr Ile Leu Leu Phe Ala Ala Gly Ser Leu Ala Ile Pro
1               5                   10                  15

Ala Pro Ser Ile Arg Leu Ala Pro Pro Tyr Pro Ser Ser Gln Glu Asp
            20                  25                  30

Pro Ile His Ile Ala Cys Met Ala Pro Gly Asn Phe Pro Gly Ala Asn
        35                  40                  45

Phe Thr Leu Tyr Arg Gly Gly Gln Val Val Gln Leu Leu Gln Ala Pro
    50                  55                  60

Thr Asp Gln Arg Gly Val Thr Phe Asn Leu Ser Gly Gly Ser Ser Lys
65                  70                  75                  80

Ala Pro Gly Gly Pro Phe His Cys Gln Tyr Gly Val Leu Gly Glu Leu
                85                  90                  95

Asn Gln Ser Gln Leu Ser Asp Leu Ser Glu Pro Val Asn Val Ser Phe
            100                 105                 110

Pro Val Pro Thr Trp Ile Leu Val Leu Ser Leu Ser Leu Ala Gly Ala
        115                 120                 125

Leu Phe Leu Leu Ala Gly Leu Val Ala Val Ala Leu Val Val Arg Lys
    130                 135                 140

Val Lys Leu Arg Asn Leu Gln Lys Lys Arg Asp Arg Glu Ser Cys Trp
145                 150                 155                 160

Ala Gln Ile Asn Phe Asp Ser Thr Asp Met Ser Phe Asp Asn Ser Leu
                165                 170                 175

Phe Thr Val Ser Ala Lys Thr Met Pro Glu Glu Asp Pro Ala Thr Leu
            180                 185                 190

Asp Asp His Ser Gly Thr Thr Ala Thr Pro Ser Asn Ser Arg Thr Arg
        195                 200                 205

Lys Arg Pro Thr Ser Thr Ser Ser Ser Pro Glu Thr Pro Glu Phe Ser
    210                 215                 220

Thr Phe Arg Ala Cys Gln
225                 230
```

<210> 146
<211> 693
<212> DNA
<213> Homo sapiens

<400> 146
atgccctgga ccatcttgct ctttgcagct ggctccttgg cgatcccagc accatccatc 60
cggctggtgc cccgttccc aagcagccaa gaggacccca tccacatcgc atgcatggcc 120
cctgggaact tcccgggggc gaatttcaca ctgtatcgag gggggcaggt ggtccagctc 180
ctgcaggccc ccacggacca gcgcggggtg acatttaacc tgagcggcgg cagcagcaag 240
gctccagggg gacccttcca ctgccagtat ggagtgttag gtgagctcaa ccagtcccag 300
ctgtcagacc tcagcgagcc cgtgaacgtc tccttcccag tgcccacttg gatcttggtg 360
ctctccctga gcctggctgg tgccctcttc ctccttgctg ggctggtggc tgttgccctg 420
gtggtcagaa aagttaaact cagaaattta cagaagaaaa gagatcgaga atcctgctgg 480
gcccagatta acttcgacag cacagacatg tccttcgata actccctgtt taccgtctcc 540
gcgaaaacga tgccagaaga agacccggcc accttggatg atcactcagg caccactgcc 600
acccccagca actccaggac ccggaagagg cccacttcca cgtcctcctc gcctgagacc 660
cccgaattca gcactttccg ggcctgccag tga 693

<210> 147
<211> 230
<212> PRT
<213> Homo sapiens

<400> 147
Met Pro Trp Thr Ile Leu Leu Phe Ala Ala Gly Ser Leu Ala Ile Pro
1 5 10 15

Ala Pro Ser Ile Arg Leu Val Pro Pro Phe Pro Ser Ser Gln Glu Asp
20 25 30

Pro Ile His Ile Ala Cys Met Ala Pro Gly Asn Phe Pro Gly Ala Asn
35 40 45

Phe Thr Leu Tyr Arg Gly Gly Gln Val Val Gln Leu Leu Gln Ala Pro
50 55 60

Thr Asp Gln Arg Gly Val Thr Phe Asn Leu Ser Gly Gly Ser Ser Lys
65 70 75 80

Ala Pro Gly Gly Pro Phe His Cys Gln Tyr Gly Val Leu Gly Glu Leu
85 90 95

Asn Gln Ser Gln Leu Ser Asp Leu Ser Glu Pro Val Asn Val Ser Phe
100 105 110

Pro Val Pro Thr Trp Ile Leu Val Leu Ser Leu Ser Leu Ala Gly Ala
115 120 125

Leu Phe Leu Leu Ala Gly Leu Val Ala Val Ala Leu Val Val Arg Lys
130 135 140

Val Lys Leu Arg Asn Leu Gln Lys Lys Arg Asp Arg Glu Ser Cys Trp
145 150 155 160

Ala Gln Ile Asn Phe Asp Ser Thr Asp Met Ser Phe Asp Asn Ser Leu
165 170 175

Phe Thr Val Ser Ala Lys Thr Met Pro Glu Glu Asp Pro Ala Thr Leu
180 185 190

Asp Asp His Ser Gly Thr Thr Ala Thr Pro Ser Asn Ser Arg Thr Arg
195 200 205

Lys Arg Pro Thr Ser Thr Ser Ser Ser Pro Glu Thr Pro Glu Phe Ser
210 215 220

Thr Phe Arg Ala Cys Gln
225 230

<210> 148
<211> 693
<212> DNA
<213> Homo sapiens

<400> 148
```
atgccctgga ccatcttgct ctttgcagct ggctccttgg cgatcccagc accatccatc 60
cggctggtgc ccccgtaccc aagcagccaa gaggacccca tccacatcgc atgcatggcc 120
cctgggaact tcccgggggc gaatttcaca ctgtatcgag gggggcaggt ggtccagctc 180
ctgcaggccc ccacggacca gcacggggtg acatttaacc tgagcggcgg cagcagcaag 240
gctccagggg gacccttcca ctgccagtat ggagtgttag gtgagctcaa ccagtcccag 300
ctgtcagacc tcagcgagcc cgtgaacgtc tccttcccag tgcccacttg gatcttggtg 360
ctctccctga gcctggctgg tgccctcttc ctccttgctg ggctggtggc tgttgccctg 420
gtggtcagaa aagttaaact cagaaattta cagaagaaaa gagatcgaga atcctgctgg 480
gcccagatta acttcgacag cacagacatg tccttcgata actccctgtt taccgtctcc 540
gcgaaaacga tgccagaaga agacccggcc accttggatg atcactcagg caccactgcc 600
acccccagca actccaggac ccggaagagg cccacttcca cgtcctcctc gcctgagacc 660
cccgaattca gcactttccg ggcctgccag tga                              693
```

<210> 149
<211> 230
<212> PRT
<213> Homo sapiens

<400> 149
Met Pro Trp Thr Ile Leu Leu Phe Ala Ala Gly Ser Leu Ala Ile Pro
1 5 10 15

Ala Pro Ser Ile Arg Leu Val Pro Pro Tyr Pro Ser Ser Gln Glu Asp
20 25 30

Pro Ile His Ile Ala Cys Met Ala Pro Gly Asn Phe Pro Gly Ala Asn
35 40 45

Phe Thr Leu Tyr Arg Gly Gly Gln Val Val Gln Leu Leu Gln Ala Pro
50 55 60

Thr Asp Gln His Gly Val Thr Phe Asn Leu Ser Gly Gly Ser Ser Lys
65 70 75 80

Ala Pro Gly Gly Pro Phe His Cys Gln Tyr Gly Val Leu Gly Glu Leu
85 90 95

Asn Gln Ser Gln Leu Ser Asp Leu Ser Glu Pro Val Asn Val Ser Phe
100 105 110

Pro Val Pro Thr Trp Ile Leu Val Leu Ser Leu Ser Leu Ala Gly Ala
115 120 125

Leu Phe Leu Leu Ala Gly Leu Val Ala Val Ala Leu Val Val Arg Lys
130 135 140

Val Lys Leu Arg Asn Leu Gln Lys Lys Arg Asp Arg Glu Ser Cys Trp
145 150 155 160

Ala Gln Ile Asn Phe Asp Ser Thr Asp Met Ser Phe Asp Asn Ser Leu
165 170 175

Phe Thr Val Ser Ala Lys Thr Met Pro Glu Glu Asp Pro Ala Thr Leu
180 185 190

Asp Asp His Ser Gly Thr Thr Ala Thr Pro Ser Asn Ser Arg Thr Arg
195 200 205

Lys Arg Pro Thr Ser Thr Ser Ser Ser Pro Glu Thr Pro Glu Phe Ser
210 215 220

Thr Phe Arg Ala Cys Gln
225 230

<210> 150
<211> 693
<212> DNA
<213> Homo sapiens

<400> 150
atgccctgga ccatcttgct ctttgcagct ggctccttgg cgatcccagc accatccatc 60
cggctggtgc cccgtaccc aagcagccaa gaggacccca tccacatcgc atgcatggcc 120
cctgggaact tcccgggggc gaatttcaca ctgtatcgag gggggcaggt ggtccagctc 180
ctgcaggccc ccacggacca gcgcggggtg acatttaacc tgagcggcgg cagcagcaag 240
gctccagggg gacccttcca ctgccagtat ggagtgttag gtgagctcaa ccagtcccag 300
ctgtcagagc tcagcgagcc cgtgaacgtc tccttcccag tgcccacttg gatcttggtg 360
ctctccctga gcctggctgg tgccctcttc ctccttgctg ggctggtggc tgttgccctg 420
gtggtcagaa aagttaaact cagaaattta cagaagaaaa gagatcgaga atcctgctgg 480
gcccagatta acttcgacag cacagacatg tccttcgata actccctgtt taccgtctcc 540
gcgaaaacga tgccagaaga agacccggcc accttggatg atcactcagg caccactgcc 600
acccccagca actccaggac ccggaagagg cccacttcca cgtcctcctc gcctgagacc 660
cccgaattca gcactttccg ggcctgccag tga 693

<210> 151
<211> 230
<212> PRT
<213> Homo sapiens

<400> 151
Met Pro Trp Thr Ile Leu Leu Phe Ala Ala Gly Ser Leu Ala Ile Pro
1 5 10 15

Ala Pro Ser Ile Arg Leu Val Pro Pro Tyr Pro Ser Ser Gln Glu Asp
20 25 30

Pro Ile His Ile Ala Cys Met Ala Pro Gly Asn Phe Pro Gly Ala Asn
35 40 45

Phe Thr Leu Tyr Arg Gly Gly Gln Val Val Gln Leu Leu Gln Ala Pro
50 55 60

Thr Asp Gln Arg Gly Val Thr Phe Asn Leu Ser Gly Gly Ser Ser Lys
65 70 75 80

Ala Pro Gly Gly Pro Phe His Cys Gln Tyr Gly Val Leu Gly Glu Leu
85 90 95

Asn Gln Ser Gln Leu Ser Glu Leu Ser Glu Pro Val Asn Val Ser Phe
100 105 110

Pro Val Pro Thr Trp Ile Leu Val Leu Ser Leu Ser Leu Ala Gly Ala
115 120 125

Leu Phe Leu Leu Ala Gly Leu Val Ala Val Ala Leu Val Val Arg Lys
130 135 140

Val Lys Leu Arg Asn Leu Gln Lys Lys Arg Asp Arg Glu Ser Cys Trp
145 150 155 160

Ala Gln Ile Asn Phe Asp Ser Thr Asp Met Ser Phe Asp Asn Ser Leu
165 170 175

Phe Thr Val Ser Ala Lys Thr Met Pro Glu Glu Asp Pro Ala Thr Leu
180 185 190

Asp Asp His Ser Gly Thr Thr Ala Thr Pro Ser Asn Ser Arg Thr Arg
195 200 205

Lys Arg Pro Thr Ser Thr Ser Ser Ser Pro Glu Thr Pro Glu Phe Ser
210 215 220

Thr Phe Arg Ala Cys Gln
225 230

<210> 152
<211> 249
<212> DNA
<213> Homo sapiens

<400> 152
atgtataaac tatacatacg tacatacata tgtgtttata catacacaat gcctataatg 60
attcttcact taatttttca aatttctcat caagtattgg tcttaattgt tccttttaag 120
agtgcttctg taagtattaa atctaactta tatattccat taatttgtaa tttaattgcg 180
tgtccaatgt acagcagtaa caatcagaat cttcacaaag gccagtgcca ttttgtaaaa 240
t[illegible]tttttaa 249

<210> 153
<211> 82
<212> PRT
<213> Homo sapiens

<400> 153
Met Tyr Lys Leu Tyr Ile Arg Thr Tyr Ile Cys Val Tyr Thr Tyr Thr
1 5 10 15

Met Pro Ile Met Ile Leu His Leu Ile Phe Gln Ile Ser His Gln Val
20 25 30

Leu Val Leu Ile Val Pro Phe Lys Ser Ala Ser Val Ser Ile Lys Ser
35 40 45

Asn Leu Tyr Ile Pro Leu Ile Cys Asn Leu Ile Ala Cys Pro Met Tyr
50 55 60

Ser Ser Asn Asn Gln Asn Leu His Lys Gly Gln Cys His Phe Val Lys
65 70 75 80

Ser Phe

<210> 154
<211> 249
<212> DNA
<213> Homo sapiens

<400> 154
atgtataaac tatacataca tacatacata tgtgcttata catacacaat gcctataatg 60
attcttcact taatttttca aatttctcat caagtattgg tcttaattgt tccttttaag 120
agtgcttctg taagtattaa atctaactta tatattccat taatttgtaa tttaattgcg 180
tgtccaatgt acagcagtaa caatcagaat cttcacaaag gccagtgcca ttttgtaaaa 240
tctttttaa 249

<210> 155
<211> 82
<212> PRT
<213> Homo sapiens

<400> 155
Met Tyr Lys Leu Tyr Ile His Thr Tyr Ile Cys Ala Tyr Thr Tyr Thr
1 5 10 15

Met Pro Ile Met Ile Leu His Leu Ile Phe Gln Ile Ser His Gln Val
20 25 30

Leu Val Leu Ile Val Pro Phe Lys Ser Ala Ser Val Ser Ile Lys Ser
35 40 45

Asn Leu Tyr Ile Pro Leu Ile Cys Asn Leu Ile Ala Cys Pro Met Tyr
50 55 60

Ser Ser Asn Asn Gln Asn Leu His Lys Gly Gln Cys His Phe Val Lys
65 70 75 80

Ser Phe

<210> 156
<211> 249
<212> DNA
<213> Homo sapiens

<400> 156
atgtataaac tatacataca tacatacata tgtgtttata catacacaat gcctataatg 60
attcttcact taatttttca aattactcat caagtattgg tcttaattgt tccttttaag 120
agtgcttctg taagtattaa atctaactta tatattccat taatttgtaa tttaattgcg 180
tgtccaatgt acagcagtaa caatcagaat cttcacaaag gccagtgcca ttttgtaaaa 240
tctttttaa 249

<210> 157
<211> 82
<212> PRT
<213> Homo sapiens

<400> 157
Met Tyr Lys Leu Tyr Ile His Thr Tyr Ile Cys Val Tyr Thr Tyr Thr
1 5 10 15

Met Pro Ile Met Ile Leu His Leu Ile Phe Gln Ile Thr His Gln Val
20 25 30

Leu Val Leu Ile Val Pro Phe Lys Ser Ala Ser Val Ser Ile Lys Ser
35 40 45

Asn Leu Tyr Ile Pro Leu Ile Cys Asn Leu Ile Ala Cys Pro Met Tyr
50 55 60

Ser Ser Asn Asn Gln Asn Leu His Lys Gly Gln Cys His Phe Val Lys
65 70 75 80

Ser Phe

<210> 158
<211> 249
<212> DNA
<213> Homo sapiens

<400> 158
atgtataaac tatacataca tacatacata tgtgtttata catacacaat gcctataatg 60
attcttcact taatttttca aatttctcat gaagtattgg tcttaattgt tccttttaag 120
agtgcttctg taagtattaa atctaactta tatattccat taatttgtaa tttaattgcg 180
tgtccaatgt acagcagtaa caatcagaat cttcacaaag gccagtgcca ttttgtaaaa 240
tctttttaa 249

<210> 159
<211> 82
<212> PRT
<213> Homo sapiens

<400> 159
Met Tyr Lys Leu Tyr Ile His Thr Tyr Ile Cys Val Tyr Thr Tyr Thr
1 5 10 15

Met Pro Ile Met Ile Leu His Leu Ile Phe Gln Ile Ser His Glu Val
20 25 30

Leu Val Leu Ile Val Pro Phe Lys Ser Ala Ser Val Ser Ile Lys Ser
35 40 45

Asn Leu Tyr Ile Pro Leu Ile Cys Asn Leu Ile Ala Cys Pro Met Tyr
50 55 60

Ser Ser Asn Asn Gln Asn Leu His Lys Gly Gln Cys His Phe Val Lys
65 70 75 80

Ser Phe

<210> 160
<211> 498
<212> DNA
<213> Homo sapiens

<400> 160
atgctggtgg cagtcctggc atgccaccgg ggggcacggc gccccatgcc aggcggcact 60
cgctgccgag tcctactgct cagtctcacc tttggcacgt ccatggcctg cggcaacgtg 120
ggcctaaggg ctgtgcccct ggacctggca caactggtta ctaccaccac acctctgttc 180
accctggccc tgtcggcgct gctgctgggc cgccgccacc acccgcttca gttggccgcc 240
atgggtccgc tctgcctggg ggccgcctgc agcctggctg gagagttccg gacaccccct 300
accggctgtg gcttcctgct cgcagccacc tgcctccgcg gactcaagtc ggttcagcaa 360
aacagggtct ggctctgtca cccaggctgc attggtgaga tctcagctca atacagcctc 420
cgcatcctgg gttcaagtga ttcttctgcc tcagcctccc aagtgccctg ctgcaggagg 480
agaggctgga cgcggtga 498

<210> 161
<211> 165
<212> PRT
<213> Homo sapiens

<400> 161
Met Leu Val Ala Val Leu Ala Cys His Arg Gly Ala Arg Arg Pro Met
1 5 10 15

Pro Gly Gly Thr Arg Cys Arg Val Leu Leu Leu Ser Leu Thr Phe Gly
20 25 30

Thr Ser Met Ala Cys Gly Asn Val Gly Leu Arg Ala Val Pro Leu Asp
35 40 45

Leu Ala Gln Leu Val Thr Thr Thr Thr Pro Leu Phe Thr Leu Ala Leu
50 55 60

Ser Ala Leu Leu Leu Gly Arg Arg His His Pro Leu Gln Leu Ala Ala
65 70 75 80

Met Gly Pro Leu Cys Leu Gly Ala Ala Cys Ser Leu Ala Gly Glu Phe
85 90 95

Arg Thr Pro Pro Thr Gly Cys Gly Phe Leu Leu Ala Ala Thr Cys Leu
100 105 110

Arg Gly Leu Lys Ser Val Gln Gln Asn Arg Val Trp Leu Cys His Pro
115 120 125

Gly Cys Ile Gly Glu Ile Ser Ala Gln Tyr Ser Leu Arg Ile Leu Gly
130 135 140

Ser Ser Asp Ser Ser Ala Ser Ala Ser Gln Val Pro Cys Cys Arg Arg
145 150 155 160

Arg Gly Trp Thr Arg
165

<210> 162
<211> 498
<212> DNA
<213> Homo sapiens

<400> 162
atgctggtgg cagccctggc atgccaccgg ggggcacggc accccatgcc aggcggcact 60
cgctgccgag tcctactgct cagtctcacc tttggcacgt ccatggcctg cggcaacgtg 120
ggcctaaggg ctgtgcccct ggacctggca caactggtta ctaccaccac acctctgttc 180
accctggccc tgtcggcgct gctgctgggc cgccgccacc acccgcttca gttggccgcc 240
atgggtccgc tctgcctggg ggccgcctgc agcctggctg gagagttccg gacacccccct 300
accggctgtg gcttcctgct cgcagccacc tgcctccgcg gactcaagtc ggttcagcaa 360
aacagggtct ggctctgtca cccaggctgc attggtgaga tctcagctca atacagcctc 420
cgcatcctgg gttcaagtga ttcttctgcc tcagcctccc aagtgccctg ctgcaggagg 480
agaggctgga cgcggtga 498

<210> 163
<211> 165
<212> PRT
<213> Homo sapiens

<400> 163
Met Leu Val Ala Ala Leu Ala Cys His Arg Gly Ala Arg His Pro Met
1 5 10 15

Pro Gly Gly Thr Arg Cys Arg Val Leu Leu Leu Ser Leu Thr Phe Gly
20 25 30

Thr Ser Met Ala Cys Gly Asn Val Gly Leu Arg Ala Val Pro Leu Asp
35 40 45

Leu Ala Gln Leu Val Thr Thr Thr Thr Pro Leu Phe Thr Leu Ala Leu
50 55 60

Ser Ala Leu Leu Leu Gly Arg Arg His His Pro Leu Gln Leu Ala Ala
65 70 75 80

Met Gly Pro Leu Cys Leu Gly Ala Ala Cys Ser Leu Ala Gly Glu Phe
85 90 95

Arg Thr Pro Pro Thr Gly Cys Gly Phe Leu Leu Ala Ala Thr Cys Leu
100 105 110

Arg Gly Leu Lys Ser Val Gln Gln Asn Arg Val Trp Leu Cys His Pro
115 120 125

Gly Cys Ile Gly Glu Ile Ser Ala Gln Tyr Ser Leu Arg Ile Leu Gly
130 135 140

Ser Ser Asp Ser Ser Ala Ser Ala Ser Gln Val Pro Cys Cys Arg Arg
145 150 155 160

```
Arg Gly Trp Thr Arg
                165


<210> 164
<211> 498
<212> DNA
<213> Homo sapiens

<400> 164
atgctggtgg cagccctggc atgccaccgg ggggcacggc gccccatgcc ag[illegible]cggcagt 60
cgctgccgag tcctactgct cagtctcacc tttggcacgt ccatggcctg cggcaacgtg 120
ggcctaaggg ctgtgcccct ggacctggca caactggtta ctaccaccac acctctgttc 180
accctggccc tgtcggcgct gctgctgggc cgccgccacc acccgcttca gttggccgcc 240
atgggtccgc tctgcctggg ggccgcctgc agcctggctg gagagttccg gacacccccct 300
accggctgtg gcttcctgct cgcagccacc tgcctccgcg gactcaagtc ggttcagcaa 360
aacagggtct ggctctgtca cccaggctgc attggtgaga tctcagctca atacagcctc 420
cgcatcctgg gttcaagtga ttcttctgcc tcagcctccc aagtgccctg ctgcaggagg 480
agaggctgga cgcggtga                                               498

<210> 165
<211> 165
<212> PRT
<213> Homo sapiens

<400> 165
Met Leu Val Ala Ala Leu Ala Cys His Arg Gly Ala Arg Arg Pro Met
1               5                   10                  15

Pro Gly Gly Ser Arg Cys Arg Val Leu Leu Leu Ser Leu Thr Phe Gly
            20                  25                  30

Thr Ser Met Ala Cys Gly Asn Val Gly Leu Arg Ala Val Pro Leu Asp
        35                  40                  45

Leu Ala Gln Leu Val Thr Thr Thr Thr Pro Leu Phe Thr Leu Ala Leu
    50                  55                  60

Ser Ala Leu Leu Leu Gly Arg Arg His His Pro Leu Gln Leu Ala Ala
65                  70                  75                  80

Met Gly Pro Leu Cys Leu Gly Ala Ala Cys Ser Leu Ala Gly Glu Phe
                85                  90                  95

Arg Thr Pro Pro Thr Gly Cys Gly Phe Leu Leu Ala Ala Thr Cys Leu
            100                 105                 110

Arg Gly Leu Lys Ser Val Gln Gln Asn Arg Val Trp Leu Cys His Pro
        115                 120                 125

Gly Cys Ile Gly Glu Ile Ser Ala Gln Tyr Ser Leu Arg Ile Leu Gly
    130                 135                 140

Ser Ser Asp Ser Ser Ala Ser Ala Ser Gln Val Pro Cys Cys Arg Arg
145                 150                 155                 160

Arg Gly Trp Thr Arg
                165
```

<210> 166
<211> 498
<212> DNA
<213> Homo sapiens

<400> 166
atgctggtgg cagccctggc atgccaccgg ggggcacggc gccccatgcc aggcggcact 60
cgctgccgag tcctactgct cagtctcacc tttggcacgt ccatggcctg cggcgacgtg 120
ggcctaaggg ctgtgcccct ggacctggca caactggtta ctaccaccac acctctgttc 180
accctggccc tgtcggcgct gctgctgggc cgccgccacc acccgcttca gttggccgcc 240
atgggtccgc tctgcctggg ggccgcctgc agcctggctg gagagttccg gacacccct 300
accggctgtg gcttcctgct cgcagccacc tgcctccgcg gactcaagtc ggttcagcaa 360
aacagggtct ggctctgtca cccaggctgc attggtgaga tctcagctca atacagcctc 420
cgcatcctgg gttcaagtga ttcttctgcc tcagcctccc aagtgccctg ctgcaggagg 480
agaggctgga cgcggtga 498

<210> 167
<211> 165
<212> PRT
<213> Homo sapiens

<400> 167
Met Leu Val Ala Ala Leu Ala Cys His Arg Gly Ala Arg Arg Pro Met
1 5 10 15

Pro Gly Gly Thr Arg Cys Arg Val Leu Leu Leu Ser Leu Thr Phe Gly
20 25 30

Thr Ser Met Ala Cys Gly Asp Val Gly Leu Arg Ala Val Pro Leu Asp
35 40 45

Leu Ala Gln Leu Val Thr Thr Thr Thr Pro Leu Phe Thr Leu Ala Leu
50 55 60

Ser Ala Leu Leu Leu Gly Arg Arg His His Pro Leu Gln Leu Ala Ala
65 70 75 80

Met Gly Pro Leu Cys Leu Gly Ala Ala Cys Ser Leu Ala Gly Glu Phe
85 90 95

Arg Thr Pro Pro Thr Gly Cys Gly Phe Leu Leu Ala Ala Thr Cys Leu
100 105 110

Arg Gly Leu Lys Ser Val Gln Gln Asn Arg Val Trp Leu Cys His Pro
115 120 125

Gly Cys Ile Gly Glu Ile Ser Ala Gln Tyr Ser Leu Arg Ile Leu Gly
130 135 140

Ser Ser Asp Ser Ser Ala Ser Ala Ser Gln Val Pro Cys Cys Arg Arg
145 150 155 160

Arg Gly Trp Thr Arg
165

<210> 168

<211> 180
<212> DNA
<213> Homo sapiens

<400> 168
atgctcagtc agccccttct gctctctctt cttctctact gtgcatgtcg gcttgtactt 60
ttgccagttt ctctaaagac acaaccagag gtggggtggc tgtgtgtgca caacttcaac 120
tttacatgtg gggctgagtc cctatgttgt atatccttgt gcaaaagcac aatatgttaa 180

<210> 169
<211> 59
<212> PRT
<213> Homo sapiens

<400> 169
Met Leu Ser Gln Pro Leu Leu Leu Ser Leu Leu Leu Tyr Cys Ala Cys
1 5 10 15

Arg Leu Val Leu Leu Pro Val Ser Leu Lys Thr Gln Pro Glu Val Gly
20 25 30

Trp Leu Cys Val His Asn Phe Asn Phe Thr Cys Gly Ala Glu Ser Leu
35 40 45

Cys Cys Ile Ser Leu Cys Lys Ser Thr Ile Cys
50 55

<210> 170
<211> 180
<212> DNA
<213> Homo sapiens

<400> 170
atgctcactg agccccttct gctctctctt cttctctact gtgcatgtcg gcttgtactt 60
ttgccagttt ctctaaagac acaaccagag gtggggtggc tgtgtgtgca caacttcaac 120
tttacatgtg gggctgagtc cctatgttgt atatccttgt gcaaaagcac aatatgttaa 180

<210> 171
<211> 59
<212> PRT
<213> Homo sapiens

<400> 171
Met Leu Thr Glu Pro Leu Leu Leu Ser Leu Leu Leu Tyr Cys Ala Cys
1 5 10 15

Arg Leu Val Leu Leu Pro Val Ser Leu Lys Thr Gln Pro Glu Val Gly
20 25 30

Trp Leu Cys Val His Asn Phe Asn Phe Thr Cys Gly Ala Glu Ser Leu
35 40 45

Cys Cys Ile Ser Leu Cys Lys Ser Thr Ile Cys
50 55

<210> 172

<211> 180
<212> DNA
<213> Homo sapiens

<400> 172
atgctcactc agcccgttct gctctctctt cttctctact gtgcatgtcg gcttgtactt 60
ttgccagttt ctctaaagac acaaccagag gtggggtggc tgtgtgtgca caacttcaac 120
tttacatgtg ggctgagtc cctatgttgt atatccttgt gcaaaagcac aatatgttaa 180

<210> 173
<211> 59
<212> PRT
<213> Homo sapiens

<400> 173
Met Leu Thr Gln Pro Val Leu Leu Ser Leu Leu Leu Tyr Cys Ala Cys
1 5 10 15

Arg Leu Val Leu Leu Pro Val Ser Leu Lys Thr Gln Pro Glu Val Gly
20 25 30

Trp Leu Cys Val His Asn Phe Asn Phe Thr Cys Gly Ala Glu Ser Leu
35 40 45

Cys Cys Ile Ser Leu Cys Lys Ser Thr Ile Cys
50 55

<210> 174
<211> 180
<212> DNA
<213> Homo sapiens

<400> 174
atgctcactc agccccttct gctctctctt cttctctact gtgcatgtcg gcttgtactt 60
ttgccagttt ctctaaagac acaaccagag gtggggtggc tgtgtgtgcg caacttcaac 120
tttacatgtg ggctgagtc cctatgttgt atatccttgt gcaaaagcac aatatgttaa 180

<210> 175
<211> 59
<212> PRT
<213> Homo sapiens

<400> 175
Met Leu Thr Gln Pro Leu Leu Leu Ser Leu Leu Leu Tyr Cys Ala Cys
1 5 10 15

Arg Leu Val Leu Leu Pro Val Ser Leu Lys Thr Gln Pro Glu Val Gly
20 25 30

Trp Leu Cys Val Arg Asn Phe Asn Phe Thr Cys Gly Ala Glu Ser Leu
35 40 45

Cys Cys Ile Ser Leu Cys Lys Ser Thr Ile Cys
50 55

<210> 176

<211> 177
<212> DNA
<213> Homo sapiens

<400> 176
atgagtattt tggttagggt tcacttgtac cttttaggtt tagctcttat gcaaagcgtg 60
tggttcagat ccatgtgtca ccctcaggtc actacgtccc actgcagcag gtatggggag 120
aatcataacc ataacacctt cccttgcagt gaatttctct ctcatatttg tctttag 177

<210> 177
<211> 58
<212> PRT
<213> Homo sapiens

<400> 177
Met Ser Ile Leu Val Arg Val His Leu Tyr Leu Leu Gly Leu Ala Leu
1 5 10 15

Met Gln Ser Val Trp Phe Arg Ser Met Cys His Pro Gln Val Thr Thr
20 25 30

Ser His Cys Ser Arg Tyr Gly Glu Asn His Asn His Asn Thr Phe Pro
35 40 45

Cys Ser Glu Phe Leu Ser His Ile Cys Leu
50 55

<210> 178
<211> 177
<212> DNA
<213> Homo sapiens

<400> 178
atgagtattt tggttagggt tcacttgtac cttttaggtt tagctcttat gcaaagcctg 60
tggttcaaat ccatgtgtca ccctcaggtc actacgtccc actgcagcag gtatggggag 120
aatcataacc ataacacctt cccttgcagt gaatttctct ctcatatttg tctttag 177

<210> 179
<211> 58
<212> PRT
<213> Homo sapiens

<400> 179
Met Ser Ile Leu Val Arg Val His Leu Tyr Leu Leu Gly Leu Ala Leu
1 5 10 15

Met Gln Ser Leu Trp Phe Lys Ser Met Cys His Pro Gln Val Thr Thr
20 25 30

Ser His Cys Ser Arg Tyr Gly Glu Asn His Asn His Asn Thr Phe Pro
35 40 45

Cys Ser Glu Phe Leu Ser His Ile Cys Leu
50 55

<210> 180

<211> 177
<212> DNA
<213> Homo sapiens

<400> 180
atgagtattt tggttagggt tcacttgtac cttttaggtt tagctcttat gcaaagcctg 60
tggttcagaa ccatgtgtca ccctcaggtc actacgtccc actgcagcag gtatggggag 120
aatcataacc ataacacctt cccttgcagt gaatttctct ctcatatttg tctttag 177

<210> 181
<211> 58
<212> PRT
<213> Homo sapiens

<400> 181
Met Ser Ile Leu Val Arg Val His Leu Tyr Leu Leu Gly Leu Ala Leu
1 5 10 15

Met Gln Ser Leu Trp Phe Arg Thr Met Cys His Pro Gln Val Thr Thr
20 25 30

Ser His Cys Ser Arg Tyr Gly Glu Asn His Asn His Asn Thr Phe Pro
35 40 45

Cys Ser Glu Phe Leu Ser His Ile Cys Leu
50 55

<210> 182
<211> 177
<212> DNA
<213> Homo sapiens

<400> 182
atgagtattt tggttagggt tcacttgtac cttttaggtt tagctcttat gcaaagcctg 60
tggttcagat ccatgtgtca ccctcaggtc actacgtccc actgcagcag gtatggggac 120
aatcataacc ataacacctt cccttgcagt gaatttctct ctcatatttg tctttag 177

<210> 183
<211> 58
<212> PRT
<213> Homo sapiens

<400> 183
Met Ser Ile Leu Val Arg Val His Leu Tyr Leu Leu Gly Leu Ala Leu
1 5 10 15

Met Gln Ser Leu Trp Phe Arg Ser Met Cys His Pro Gln Val Thr Thr
20 25 30

Ser His Cys Ser Arg Tyr Gly Asp Asn His Asn His Asn Thr Phe Pro
35 40 45

Cys Ser Glu Phe Leu Ser His Ile Cys Leu
50 55

<210> 184

<211> 423
<212> DNA
<213> Homo sapiens

<400> 184
```
atgatcagtg gatggaagct gcctattatt attgtcgttg ttgttgtttg ccatgactgc 60
tctgggccgg gggtagagct agcatccggg catgtacgag ggaagaggga ggcaggcctc 120
tattcaaagg cagaaattcc tttaagattg tggtctgctg ggtttcaggg agtgtctgtg 180
ttgtttgttt ttgtttgttt gtttgttttg agacagggtc tcgctctgtc acccaggctg 240
gagtgtagtg gtgcagtctt ggctcactgc aacctccacc tcctgggctc aagcgattct 300
catgcctcag cctcccgagt agctgggact acaggtgtgt gccactatgc ctggctaatt 360
tttgtatttt ttgtagagac ggggttttgc catgttgccc aggctggaag tgtctatgtt 420
taa                                                               423
```

<210> 185
<211> 140
<212> PRT
<213> Homo sapiens

<400> 185
```
Met Ile Ser Gly Trp Lys Leu Pro Ile Ile Ile Val Val Val Val Val
1               5                   10                  15

Cys His Asp Cys Ser Gly Pro Gly Val Glu Leu Ala Ser Gly His Val
            20                  25                  30

Arg Gly Lys Arg Glu Ala Gly Leu Tyr Ser Lys Ala Glu Ile Pro Leu
        35                  40                  45

Arg Leu Trp Ser Ala Gly Phe Gln Gly Val Ser Val Leu Phe Val Phe
    50                  55                  60

Val Cys Leu Phe Val Leu Arg Gln Gly Leu Ala Leu Ser Pro Arg Leu
65                  70                  75                  80

Glu Cys Ser Gly Ala Val Leu Ala His Cys Asn Leu His Leu Leu Gly
                85                  90                  95

Ser Ser Asp Ser His Ala Ser Ala Ser Arg Val Ala Gly Thr Thr Gly
            100                 105                 110

Val Cys His Tyr Ala Trp Leu Ile Phe Val Phe Phe Val Glu Thr Gly
        115                 120                 125

Phe Cys His Val Ala Gln Ala Gly Ser Val Tyr Val
    130                 135                 140
```

<210> 186
<211> 423
<212> DNA
<213> Homo sapiens

<400> 186
```
atgctcactg gatggaagct gcctattatt attgtcgttg ttgttgtttg ccatgactgc 60
tctgggccgg gggtagagct agcatccggg catgtacgag ggaagaggga ggcaggcctc 120
tattcaaagg cagaaattcc tttaagattg tggtctgctg ggtttcaggg agtgtctgtg 180
ttgtttgttt ttgtttgttt gtttgttttg agacagggtc tcgctctgtc acccaggctg 240
```

gagtgtagtg gtgcagtctt ggctcactgc aacctccacc tcctgggctc aagcgattct 300
catgcctcag cctcccgagt agctgggact acaggtgtgt gccactatgc ctggctaatt 360
tttgtatttt ttgtagagac ggggttttgc catgttgccc aggctggaag tgtctatgtt 420
taa 423

<210> 187
<211> 140
<212> PRT
<213> Homo sapiens

<400> 187
Met Leu Thr Gly Trp Lys Leu Pro Ile Ile Ile Val Val Val Val Val
1 5 10 15

Cys His Asp Cys Ser Gly Pro Gly Val Glu Leu Ala Ser Gly His Val
20 25 30

Arg Gly Lys Arg Glu Ala Gly Leu Tyr Ser Lys Ala Glu Ile Pro Leu
35 40 45

Arg Leu Trp Ser Ala Gly Phe Gln Gly Val Ser Val Leu Phe Val Phe
50 55 60

Val Cys Leu Phe Val Leu Arg Gln Gly Leu Ala Leu Ser Pro Arg Leu
65 70 75 80

Glu Cys Ser Gly Ala Val Leu Ala His Cys Asn Leu His Leu Leu Gly
85 90 95

Ser Ser Asp Ser His Ala Ser Ala Ser Arg Val Ala Gly Thr Thr Gly
100 105 110

Val Cys His Tyr Ala Trp Leu Ile Phe Val Phe Phe Val Glu Thr Gly
115 120 125

Phe Cys His Val Ala Gln Ala Gly Ser Val Tyr Val
130 135 140

<210> 188
<211> 423
<212> DNA
<213> Homo sapiens

<400> 188
atgctcagtg gatggaggct gcctattatt attgtcgttg ttgttgtttg ccatgactgc 60
tctgggccgg gggtagagct agcatccggg catgtacgag ggaagaggga ggcaggcctc 120
tattcaaagg cagaaattcc tttaagattg tggtctgctg ggtttcaggg agtgtctgtg 180
ttgtttgttt ttgtttgttt gtttgttttg agacagggtc tcgctctgtc acccaggctg 240
gagtgtagtg gtgcagtctt ggctcactgc aacctccacc tcctgggctc aagcgattct 300
catgcctcag cctcccgagt agctgggact acaggtgtgt gccactatgc ctggctaatt 360
tttgtatttt ttgtagagac ggggttttgc catgttgccc aggctggaag tgtctatgtt 420
taa 423

<210> 189
<211> 140
<212> PRT
<213> Homo sapiens

<400> 189
Met Leu Ser Gly Trp Arg Leu Pro Ile Ile Ile Val Val Val Val Val
1 5 10 15

Cys His Asp Cys Ser Gly Pro Gly Val Glu Leu Ala Ser Gly His Val
20 25 30

Arg Gly Lys Arg Glu Ala Gly Leu Tyr Ser Lys Ala Glu Ile Pro Leu
35 40 45

Arg Leu Trp Ser Ala Gly Phe Gln Gly Val Ser Val Leu Phe Val Phe
50 55 60

Val Cys Leu Phe Val Leu Arg Gln Gly Leu Ala Leu Ser Pro Arg Leu
65 70 75 80

Glu Cys Ser Gly Ala Val Leu Ala His Cys Asn Leu His Leu Leu Gly
85 90 95

Ser Ser Asp Ser His Ala Ser Ala Ser Arg Val Ala Gly Thr Thr Gly
100 105 110

Val Cys His Tyr Ala Trp Leu Ile Phe Val Phe Phe Val Glu Thr Gly
115 120 125

Phe Cys His Val Ala Gln Ala Gly Ser Val Tyr Val
130 135 140

<210> 190
<211> 423
<212> DNA
<213> Homo sapiens

<400> 190
atgctcagtg gatggaagct gcctattatt attgtcgttg ttgttgtttg ccatgagtgc 60
tctgggccgg gggtagagct agcatccggg catgtacgag ggaagaggga ggcaggcctc 120
tattcaaagg cagaaattcc tttaagattg tggtctgctg ggtttcaggg agtgtctgtg 180
ttgtttgttt ttgtttgttt gtttgttttg agacagggtc tcgctctgtc acccaggctg 240
gagtgtagtg gtgcagtctt ggctcactgc aacctccacc tcctgggctc aagcgattct 300
catgcctcag cctcccgagt agctgggact acaggtgtgt gccactatgc ctggctaatt 360
tttgtatttt ttgtagagac ggggttttgc catgttgccc aggctggaag tgtctatgtt 420
taa 423

<210> 191
<211> 140
<212> PRT
<213> Homo sapiens

<400> 191
Met Leu Ser Gly Trp Lys Leu Pro Ile Ile Ile Val Val Val Val Val
1 5 10 15

Cys His Glu Cys Ser Gly Pro Gly Val Glu Leu Ala Ser Gly His Val
20 25 30

Arg Gly Lys Arg Glu Ala Gly Leu Tyr Ser Lys Ala Glu Ile Pro Leu
35 40 45

```
Arg Leu Trp Ser Ala Gly Phe Gln Gly Val Ser Val Leu Phe Val Phe
     50                  55                  60

Val Cys Leu Phe Val Leu Arg Gln Gly Leu Ala Leu Ser Pro Arg Leu
 65                  70                  75                  80

Glu Cys Ser Gly Ala Val Leu Ala His Cys Asn Leu His Leu Leu Gly
                 85                  90                  95

Ser Ser Asp Ser His Ala Ser Ala Ser Arg Val Ala Gly Thr Thr Gly
            100                 105                 110

Val Cys His Tyr Ala Trp Leu Ile Phe Val Phe Phe Val Glu Thr Gly
        115                 120                 125

Phe Cys His Val Ala Gln Ala Gly Ser Val Tyr Val
    130                 135                 140


<210> 192
<211> 1422
<212> DNA
<213> Homo sapiens

<400> 192
atgaggaggg cgtccgctgg agggagccgg ctgctggcat gggtgctgtg gctgcaggcc 60
tggcaggtgg cagccccatg cccaggtgcc tgcgtatgct acaatgagcc caaggtgacg 120
acaagctgcc cccagcaggg cctgcaggct gtgcccgtgg gcatccctgc tgccagccag 180
cgcatcttcc tgcacggcaa ccgcatctcg catgtgccag ctgccagctt ccgtgcctgc 240
cgcaacctca ccatcctgtg gctgcactcg aatgtgctgg cccgaattga tgcggctgcc 300
ttcactggcc tggccctcct ggagcagctg gacctcagcg ataatgcaca gctccggtct 360
gtggaccctg ccacattcca cggcctgggc cgcgtacaca cgctgcacct ggaccgctgc 420
ggcctgcagg agctgggccc ggggctgttc cgcggcctgg ctgccctgca gtacctctac 480
ctgcaggaca acgcgctgca ggcactgcct gatgacacct tccgcgacct gggcaacctc 540
acacacctct tcctgcacgg caaccgcatc tccagcgtgc ccgagcgcgc cttccgtggg 600
ctgcacagcc tcgaccgtct cctactgcac cagaaccgcg tggcccatgt gcacccgcat 660
gccttccgtg accttggccg cctcatgaca ctctatctgt ttgccaacaa tctatcagcg 720
ctgcccactg aggccctggc ccccctgcgt gccctgcagt acctgaggct caacgacaac 780
ccctgggtgt gtgactgccg ggcacgccca ctctgggcct ggctgcagaa gttccgcggc 840
tcctcctccg aggtgccctg cagcctcccg caacgcctgg ctggccgtga cctcaaacgc 900
ctagctgcca atgacctgca gggctgcgct gtggccaccg gcccttacca tcccatctgg 960
accggcaggg ccaccgatga ggagccgctg gggcttccca agtgctgcca gccagatgcc 1020
gctgacaagg cctcagtact ggagcctgga agaccagctt cggcaggcaa tgcgctgaag 1080
ggacgcgtgc cgcccggtga cagcccgccg ggcaacggct ctggcccacg gcacatcaat 1140
gactcaccct ttgggactct gcctggctct gctgagcccc cgctcactgc agtgcggccc 1200
gagggctccg agccaccagg gttccccacc tcgggccctc gccggaggcc aggctgttca 1260
cgcaagaacc gcacccgcag ccactgccgt ctgggccagg caggcagcgg gggtggcggg 1320
actggtgact cagaaggctc aggtgcccta cccagcctca cctgcagcct cacccccctg 1380
ggcctggcgc tggtgctgtg gacagtgctt gggccctgct ga                    1422

<210> 193
<211> 473
<212> PRT
<213> Homo sapiens

<400> 193
Met Arg Arg Ala Ser Ala Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
  1               5                  10                  15
```

Trp Leu Gln Ala Trp Gln Val Ala Ala Pro Cys Pro Gly Ala Cys Val
20 25 30

Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser Cys Pro Gln Gln Gly Leu
35 40 45

Gln Ala Val Pro Val Gly Ile Pro Ala Ala Ser Gln Arg Ile Phe Leu
50 55 60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Arg Ala Cys
65 70 75 80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Val Leu Ala Arg Ile
85 90 95

Asp Ala Ala Ala Phe Thr Gly Leu Ala Leu Leu Glu Gln Leu Asp Leu
100 105 110

Ser Asp Asn Ala Gln Leu Arg Ser Val Asp Pro Ala Thr Phe His Gly
115 120 125

Leu Gly Arg Val His Thr Leu His Leu Asp Arg Cys Gly Leu Gln Glu
130 135 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145 150 155 160

Leu Gln Asp Asn Ala Leu Gln Ala Leu Pro Asp Asp Thr Phe Arg Asp
165 170 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Ser Ser
180 185 190

Val Pro Glu Arg Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
195 200 205

Leu His Gln Asn Arg Val Ala His Val His Pro His Ala Phe Arg Asp
210 215 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Ala
225 230 235 240

Leu Pro Thr Glu Ala Leu Ala Pro Leu Arg Ala Leu Gln Tyr Leu Arg
245 250 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
260 265 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Ser
275 280 285

Leu Pro Gln Arg Leu Ala Gly Arg Asp Leu Lys Arg Leu Ala Ala Asn
290 295 300

Asp Leu Gln Gly Cys Ala Val Ala Thr Gly Pro Tyr His Pro Ile Trp
305 310 315 320

Thr Gly Arg Ala Thr Asp Glu Glu Pro Leu Gly Leu Pro Lys Cys Cys

```
                325                 330                 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
            340                 345                 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Ser
        355                 360                 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
    370                 375                 380

Gly Thr Leu Pro Gly Ser Ala Glu Pro Pro Leu Thr Ala Val Arg Pro
385                 390                 395                 400

Glu Gly Ser Glu Pro Pro Gly Phe Pro Thr Ser Gly Pro Arg Arg Arg
                405                 410                 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
            420                 425                 430

Gln Ala Gly Ser Gly Gly Gly Gly Thr Gly Asp Ser Glu Gly Ser Gly
        435                 440                 445

Ala Leu Pro Ser Leu Thr Cys Ser Leu Thr Pro Leu Gly Leu Ala Leu
    450                 455                 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465                 470


&lt;210&gt; 194
&lt;211&gt; 1422
&lt;212&gt; DNA
&lt;213&gt; Homo sapiens

&lt;400&gt; 194
atgaagaggg cgtccgttgg agggagccgg ctgctggcat gggtgctgtg gctgcaggcc 60
tggcaggtgg cagccccatg cccaggtgcc tgcgtatgct acaatgagcc caaggtgacg 120
acaagctgcc cccagcaggg cctgcaggct gtgcccgtgg gcatccctgc tgccagccag 180
cgcatcttcc tgcacggcaa ccgcatctcg catgtgccag ctgccagctt ccgtgcctgc 240
cgcaacctca ccatcctgtg gctgcactcg aatgtgctgg cccgaattga tgcggctgcc 300
ttcactggcc tggccctcct ggagcagctg gacctcagcg ataatgcaca gctccggtct 360
gtggaccctg ccacattcca cggcctgggc cgcgtacaca cgctgcacct ggaccgctgc 420
ggcctgcagg agctgggccc ggggctgttc cgcggcctgg ctgccctgca gtacctctac 480
ctgcaggaca acgcgctgca ggcactgcct gatgacacct tccgcgacct gggcaacctc 540
acacacctct tcctgcacgg caaccgcatc tccagcgtgc ccgagcgcgc cttccgtggg 600
ctgcacagcc tcgaccgtct cctactgcac cagaaccgcg tggcccatgt gcacccgcat 660
gccttccgtg accttggccg cctcatgaca ctctatctgt ttgccaacaa tctatcagcg 720
ctgcccactg aggccctggc ccccctgcgt gccctgcagt acctgaggct caacgacaac 780
ccctgggtgt gtgactgccg ggcacgccca ctctgggcct ggctgcagaa gttccgcggc 840
tcctcctccg aggtgccctg cagcctcccg caacgcctgg ctggccgtga cctcaaacgc 900
ctagctgcca atgacctgca gggctgcgct gtggccaccg gcccttacca tcccatctgg 960
accggcaggg ccaccgatga ggagccgctg gggcttccca agtgctgcca gccagatgcc 1020
gctgacaagg cctcagtact ggagcctgga agaccagctt cggcaggcaa tgcgctgaag 1080
ggacgcgtgc cgcccggtga cagcccgccg ggcaacggct ctggcccacg gcacatcaat 1140
gactcaccct ttgggactct gcctggctct gctgagcccc cgctcactgc agtgcggccc 1200
gagggctccg agccaccagg gttccccacc tcgggccctc gccggaggcc aggctgttca 1260
cgcaagaacc gcacccgcag ccactgccgt ctgggccagg caggcagcgg gggtggcggg 1320
```

```
actggtgact cagaaggctc aggtgcccta cccagcctca cctgcagcct caccccсctg 1380
ggcctggcgc tggtgctgtg gacagtgctt gggccctgct ga                    1422

&lt;210&gt; 195
&lt;211&gt; 473
&lt;212&gt; PRT
&lt;213&gt; Homo sapiens

&lt;400&gt; 195
Met Lys Arg Ala Ser Val Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1               5                   10                  15

Trp Leu Gln Ala Trp Gln Val Ala Ala Pro Cys Pro Gly Ala Cys Val
            20                  25                  30

Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser Cys Pro Gln Gln Gly Leu
        35                  40                  45

Gln Ala Val Pro Val Gly Ile Pro Ala Ala Ser Gln Arg Ile Phe Leu
    50                  55                  60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Arg Ala Cys
65                  70                  75                  80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Val Leu Ala Arg Ile
                85                  90                  95

Asp Ala Ala Ala Phe Thr Gly Leu Ala Leu Leu Glu Gln Leu Asp Leu
            100                 105                 110

Ser Asp Asn Ala Gln Leu Arg Ser Val Asp Pro Ala Thr Phe His Gly
        115                 120                 125

Leu Gly Arg Val His Thr Leu His Leu Asp Arg Cys Gly Leu Gln Glu
    130                 135                 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145                 150                 155                 160

Leu Gln Asp Asn Ala Leu Gln Ala Leu Pro Asp Asp Thr Phe Arg Asp
                165                 170                 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Ser Ser
            180                 185                 190

Val Pro Glu Arg Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
        195                 200                 205

Leu His Gln Asn Arg Val Ala His Val His Pro His Ala Phe Arg Asp
    210                 215                 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Ala
225                 230                 235                 240

Leu Pro Thr Glu Ala Leu Ala Pro Leu Arg Ala Leu Gln Tyr Leu Arg
                245                 250                 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
```

260 265 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Ser
275 280 285

Leu Pro Gln Arg Leu Ala Gly Arg Asp Leu Lys Arg Leu Ala Ala Asn
290 295 300

Asp Leu Gln Gly Cys Ala Val Ala Thr Gly Pro Tyr His Pro Ile Trp
305 310 315 320

Thr Gly Arg Ala Thr Asp Glu Glu Pro Leu Gly Leu Pro Lys Cys Cys
325 330 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
340 345 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Ser
355 360 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
370 375 380

Gly Thr Leu Pro Gly Ser Ala Glu Pro Pro Leu Thr Ala Val Arg Pro
385 390 395 400

Glu Gly Ser Glu Pro Pro Gly Phe Pro Thr Ser Gly Pro Arg Arg Arg
405 410 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
420 425 430

Gln Ala Gly Ser Gly Gly Gly Gly Thr Gly Asp Ser Glu Gly Ser Gly
435 440 445

Ala Leu Pro Ser Leu Thr Cys Ser Leu Thr Pro Leu Gly Leu Ala Leu
450 455 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465 470

<210> 196
<211> 1422
<212> DNA
<213> Homo sapiens

<400> 196
atgaagaggg cgtccgctgg agggagccgg ctgctggcat gggtgctgtg gctggaggcc 60
tggcaggtgg cagccccatg cccaggtgcc tgcgtatgct acaatgagcc caaggtgacg 120
acaagctgcc cccagcaggg cctgcaggct gtgcccgtgg gcatccctgc tgccagccag 180
cgcatcttcc tgcacggcaa ccgcatctcg catgtgccag ctgccagctt ccgtgcctgc 240
cgcaacctca ccatcctgtg gctgcactcg aatgtgctgg cccgaattga tgcggctgcc 300
ttcactggcc tggccctcct ggagcagctg gacctcagcg ataatgcaca gctccggtct 360
gtggaccctg ccacattcca cggcctgggc cgcgtacaca cgctgcacct ggaccgctgc 420
ggcctgcagg agctgggccc ggggctgttc cgcggcctgg ctgccctgca gtacctctac 480
ctgcaggaca acgcgctgca ggcactgcct gatgacacct tccgcgacct gggcaacctc 540
acacacctct tcctgcacgg caaccgcatc tccagcgtgc ccgagcgcgc cttccgtggg 600

ctgcacagcc tcgaccgtct cctactgcac cagaaccgcg tggcccatgt gcacccgcat 660
gccttccgtg accttggccg cctcatgaca ctctatctgt ttgccaacaa tctatcagcg 720
ctgcccactg aggccctggc ccccctgcgt gccctgcagt acctgaggct caacgacaac 780
ccctgggtgt gtgactgccg ggcacgccca ctctgggcct ggctgcagaa gttccgcggc 840
tcctcctccg aggtgccctg cagcctcccg caacgcctgg ctggccgtga cctcaaacgc 900
ctagctgcca atgacctgca gggctgcgct gtggccaccg gcccttacca tcccatctgg 960
accggcaggg ccaccgatga ggagccgctg gggcttccca agtgctgcca gccagatgcc 1020
gctgacaagg cctcagtact ggagcctgga agaccagctt cggcaggcaa tgcgctgaag 1080
ggacgcgtgc cgcccggtga cagcccgccg ggcaacggct ctggcccacg gcacatcaat 1140
gactcaccct ttgggactct gcctggctct gctgagcccc cgctcactgc agtgcggccc 1200
gagggctccg agccaccagg gttccccacc tcgggccctc gccggaggcc aggctgttca 1260
cgcaagaacc gcacccgcag ccactgccgt ctgggccagg caggcagcgg gggtggcggg 1320
actggtgact cagaaggctc aggtgcccta cccagcctca cctgcagcct caccccctg 1380
ggcctggcgc tggtgctgtg gacagtgctt gggccctgct ga 1422

<210> 197
<211> 473
<212> PRT
<213> Homo sapiens

<400> 197
Met Lys Arg Ala Ser Ala Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1 5 10 15

Trp Leu Glu Ala Trp Gln Val Ala Ala Pro Cys Pro Gly Ala Cys Val
20 25 30

Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser Cys Pro Gln Gln Gly Leu
35 40 45

Gln Ala Val Pro Val Gly Ile Pro Ala Ala Ser Gln Arg Ile Phe Leu
50 55 60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Arg Ala Cys
65 70 75 80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Val Leu Ala Arg Ile
85 90 95

Asp Ala Ala Ala Phe Thr Gly Leu Ala Leu Leu Glu Gln Leu Asp Leu
100 105 110

Ser Asp Asn Ala Gln Leu Arg Ser Val Asp Pro Ala Thr Phe His Gly
115 120 125

Leu Gly Arg Val His Thr Leu His Leu Asp Arg Cys Gly Leu Gln Glu
130 135 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145 150 155 160

Leu Gln Asp Asn Ala Leu Gln Ala Leu Pro Asp Asp Thr Phe Arg Asp
165 170 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Ser Ser
180 185 190

Val Pro Glu Arg Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu

```
        195                 200                 205

Leu His Gln Asn Arg Val Ala His Val His Pro His Ala Phe Arg Asp
    210                 215                 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Ala
225                 230                 235                 240

Leu Pro Thr Glu Ala Leu Ala Pro Leu Arg Ala Leu Gln Tyr Leu Arg
                245                 250                 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
            260                 265                 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Ser
        275                 280                 285

Leu Pro Gln Arg Leu Ala Gly Arg Asp Leu Lys Arg Leu Ala Ala Asn
    290                 295                 300

Asp Leu Gln Gly Cys Ala Val Ala Thr Gly Pro Tyr His Pro Ile Trp
305                 310                 315                 320

Thr Gly Arg Ala Thr Asp Glu Glu Pro Leu Gly Leu Pro Lys Cys Cys
                325                 330                 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
            340                 345                 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Ser
        355                 360                 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
    370                 375                 380

Gly Thr Leu Pro Gly Ser Ala Glu Pro Pro Leu Thr Ala Val Arg Pro
385                 390                 395                 400

Glu Gly Ser Glu Pro Pro Gly Phe Pro Thr Ser Gly Pro Arg Arg Arg
                405                 410                 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
            420                 425                 430

Gln Ala Gly Ser Gly Gly Gly Gly Thr Gly Asp Ser Glu Gly Ser Gly
        435                 440                 445

Ala Leu Pro Ser Leu Thr Cys Ser Leu Thr Pro Leu Gly Leu Ala Leu
    450                 455                 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465                 470


<210> 198
<211> 1422
<212> DNA
<213> Homo sapiens
```

<400> 198
atgaagaggg cgtccgctgg agggagccgg ctgctggcat gggtgctgtg gctgcaggcc 60
tggcaggtgg cagccccatg cccaggtgcc tgcgtatgct acaatgagcc caaggtgtcg 120
acaagctgcc cccagcaggg cctgcaggct gtgcccgtgg gcatccctgc tgccagccag 180
cgcatcttcc tgcacggcaa ccgcatctcg catgtgccag ctgccagctt ccgtgcctgc 240
cgcaacctca ccatcctgtg gctgcactcg aatgtgctgg cccgaattga tgcggctgcc 300
ttcactggcc tggccctcct ggagcagctg gacctcagcg ataatgcaca gctccggtct 360
gtggaccctg ccacattcca cggcctgggc cgcgtacaca cgctgcacct ggaccgctgc 420
ggcctgcagg agctgggccc ggggctgttc cgcggcctgg ctgccctgca gtacctctac 480
ctgcaggaca acgcgctgca ggcactgcct gatgacacct tccgcgacct gggcaacctc 540
acacacctct tcctgcacgg caaccgcatc tccagcgtgc ccgagcgcgc cctccgtggg 600
ctgcacagcc tcgaccgtct cctactgcac cagaaccgcg tggcccatgt gcacccgcat 660
gccttccgtg accttggccg cctcatgaca ctctatctgt ttgccaacaa tctatcagcg 720
ctgcccactg aggccctggc ccccctgcgt gccctgcagt acctgaggct caacgacaac 780
ccctgggtgt gtgactgccg ggcacgccca ctctgggcct ggctgcagaa gttccgcggc 840
tcctcctccg aggtgccctg cagcctcccg caacgcctgg ctggccgtga cctcaaacgc 900
ctagctgcca atgacctgca gggctgcgct gtggccaccg gcccttacca tcccatctgg 960
accggcaggg ccaccgatga ggagccgctg gggcttccca agtgctgcca gccagatgcc 1020
gctgacaagg cctcagtact ggagcctgga agaccagctt cggcaggcaa tgcgctgaag 1080
ggacgcgtgc cgcccggtga cagcccgccg ggcaacggct ctggcccacg gcacatcaat 1140
gactcaccct ttgggactct gcctggctct gctgagcccc cgctcactgc agtgcggccc 1200
gagggctccg agccaccagg gttccccacc tcgggccctc gccggaggcc aggctgttca 1260
cgcaagaacc gcacccgcag ccactgccgt ctgggccagg caggcagcgg gggtggcggg 1320
actggtgact cagaaggctc aggtgcccta cccagcctca cctgcagcct caccccctg 1380
ggcctggcgc tggtgctgtg gacagtgctt gggccctgct ga 1422

<210> 199
<211> 473
<212> PRT
<213> Homo sapiens

<400> 199
Met Lys Arg Ala Ser Ala Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1 5 10 15

Trp Leu Gln Ala Trp Gln Val Ala Ala Pro Cys Pro Gly Ala Cys Val
20 25 30

Cys Tyr Asn Glu Pro Lys Val Ser Thr Ser Cys Pro Gln Gln Gly Leu
35 40 45

Gln Ala Val Pro Val Gly Ile Pro Ala Ala Ser Gln Arg Ile Phe Leu
50 55 60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Arg Ala Cys
65 70 75 80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Val Leu Ala Arg Ile
85 90 95

Asp Ala Ala Ala Phe Thr Gly Leu Ala Leu Leu Glu Gln Leu Asp Leu
100 105 110

Ser Asp Asn Ala Gln Leu Arg Ser Val Asp Pro Ala Thr Phe His Gly
115 120 125

Leu Gly Arg Val His Thr Leu His Leu Asp Arg Cys Gly Leu Gln Glu
130 135 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145 150 155 160

Leu Gln Asp Asn Ala Leu Gln Ala Leu Pro Asp Asp Thr Phe Arg Asp
165 170 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Ser Ser
180 185 190

Val Pro Glu Arg Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
195 200 205

Leu His Gln Asn Arg Val Ala His Val His Pro His Ala Phe Arg Asp
210 215 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Ala
225 230 235 240

Leu Pro Thr Glu Ala Leu Ala Pro Leu Arg Ala Leu Gln Tyr Leu Arg
245 250 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
260 265 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Ser
275 280 285

Leu Pro Gln Arg Leu Ala Gly Arg Asp Leu Lys Arg Leu Ala Ala Asn
290 295 300

Asp Leu Gln Gly Cys Ala Val Ala Thr Gly Pro Tyr His Pro Ile Trp
305 310 315 320

Thr Gly Arg Ala Thr Asp Glu Glu Pro Leu Gly Leu Pro Lys Cys Cys
325 330 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
340 345 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Ser
355 360 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
370 375 380

Gly Thr Leu Pro Gly Ser Ala Glu Pro Pro Leu Thr Ala Val Arg Pro
385 390 395 400

Glu Gly Ser Glu Pro Pro Gly Phe Pro Thr Ser Gly Pro Arg Arg Arg
405 410 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
420 425 430

Gln Ala Gly Ser Gly Gly Gly Gly Thr Gly Asp Ser Glu Gly Ser Gly
435 440 445

Ala Leu Pro Ser Leu Thr Cys Ser Leu Thr Pro Leu Gly Leu Ala Leu

450 455 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465 470

<210> 200
<211> 1422
<212> DNA
<213> Mus musculus

<400> 200
atgaggaggg cgtcctccgg aggaagcagg ctgctggcat gggtgttatg gctacaggcc 60
tggagggtag caacaccatg ccctggtgct tgtgtgtgct acaatgagcc caaggtaaca 120
acaagctgcc cccagcaggg tctgcaggct gtgcccactg gcatccccagc ctctagccag 180
cgaatcttcc tgcatggcaa ccgaatctct cacgtgccag ctgcgagctt ccagtcatgc 240
cgaaatctca ctatcctgtg gctgcactct aatgcgctgg ctcggatcga tgctgctgcc 300
ttcactggtc tgaccctcct ggagcaacta gatcttagtg ataatgcaca gcttcatgtc 360
gtggacccta ccacgttcca cggcctgggc cacctgcaca cactgcacct agaccgatgt 420
ggcctgcggg agctgggtcc cggcctattc cgtggactag cagctctgca gtacctctac 480
ctacaagaca acaatctgca ggcactccct gacaacacct ttcgagacct gggcaacctc 540
acgcatctct ttctgcatgg caaccgtatc cccagtgtgc ctgagcacgc tttccgtggc 600
ctgcacagtc ttgaccgcct cctcttgcac cagaaccatg tggctcgtgt gcacccacat 660
gccttccggg accttggccg cctcatgacc ctctacctgt ttgccaacaa cctctccatg 720
ctgcctgcag aggtcctaat gcccctgagg tctctgcagt acctgcgact caatgacaac 780
ccctgggtgt gtgactgccg ggcacgtcca ctctgggcct ggctgcagaa gttccgaggt 840
tcctcatcag aggtgccctg caacctgccc caacgcctgg cagaccgtga tcttaagcgc 900
ctcgctgcca gtgacctaga gggctgtgct gtggcttcag gacccttccg tcccatccag 960
accagtcagc tcactgatga ggagctgctg agcctcccca agtgctgcca gccagatgct 1020
gcagacaaag cctcagtact ggaacccggg aggccagctt ctgccggaaa cgccctcaag 1080
ggacgtgtgc ctcccggtga cactccacca ggcaatggct caggccctcg gcacatcaat 1140
gactctccat ttggaacttt gcccagctct gcagagcccc cactgactgc cctgcggcct 1200
gggggttccg agccaccagg acttcccacc actggtcccc gcaggaggcc aggttgttcc 1260
cggaagaatc gcacccgcag ccactgccgt ctgggccagg cgggaagtgg ggccagtgga 1320
acaggggacg cagagggttc aggggctctg cctgctctgg cctgcagcct tgctcctctg 1380
ggccttgcac tggtactttg gacagtgctt gggccctgct ga 1422

<210> 201
<211> 473
<212> PRT
<213> Mus musculus

<400> 201
Met Arg Arg Ala Ser Ser Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1 5 10 15

Trp Leu Gln Ala Trp Arg Val Ala Thr Pro Cys Pro Gly Ala Cys Val
20 25 30

Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser Cys Pro Gln Gln Gly Leu
35 40 45

Gln Ala Val Pro Thr Gly Ile Pro Ala Ser Ser Gln Arg Ile Phe Leu
50 55 60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Gln Ser Cys
65 70 75 80

```
Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Ala Leu Ala Arg Ile
                85                  90                  95

Asp Ala Ala Ala Phe Thr Gly Leu Thr Leu Leu Glu Gln Leu Asp Leu
            100                 105                 110

Ser Asp Asn Ala Gln Leu His Val Val Asp Pro Thr Thr Phe His Gly
        115                 120                 125

Leu Gly His Leu His Thr Leu His Leu Asp Arg Cys Gly Leu Arg Glu
    130                 135                 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145                 150                 155                 160

Leu Gln Asp Asn Asn Leu Gln Ala Leu Pro Asp Asn Thr Phe Arg Asp
                165                 170                 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Pro Ser
            180                 185                 190

Val Pro Glu His Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
        195                 200                 205

Leu His Gln Asn His Val Ala Arg Val His Pro His Ala Phe Arg Asp
    210                 215                 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Met
225                 230                 235                 240

Leu Pro Ala Glu Val Leu Met Pro Leu Arg Ser Leu Gln Tyr Leu Arg
                245                 250                 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
            260                 265                 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Asn
        275                 280                 285

Leu Pro Gln Arg Leu Ala Asp Arg Asp Leu Lys Arg Leu Ala Ala Ser
    290                 295                 300

Asp Leu Glu Gly Cys Ala Val Ala Ser Gly Pro Phe Arg Pro Ile Gln
305                 310                 315                 320

Thr Ser Gln Leu Thr Asp Glu Glu Leu Leu Ser Leu Pro Lys Cys Cys
                325                 330                 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
            340                 345                 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Thr
        355                 360                 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
    370                 375                 380

Gly Thr Leu Pro Ser Ser Ala Glu Pro Pro Leu Thr Ala Leu Arg Pro
```

385 390 395 400

Gly Gly Ser Glu Pro Pro Gly Leu Pro Thr Thr Gly Pro Arg Arg Arg
405 410 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
420 425 430

Gln Ala Gly Ser Gly Ala Ser Gly Thr Gly Asp Ala Glu Gly Ser Gly
435 440 445

Ala Leu Pro Ala Leu Ala Cys Ser Leu Ala Pro Leu Gly Leu Ala Leu
450 455 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465 470

<210> 202
<211> 1422
<212> DNA
<213> Mus musculus

<400> 202
atgaagaggg cgtcctccgg aggaagcagg ctgctggcat gggtgttatg gctacaggtc 60
tggagggtag caacaccatg ccctggtgct tgtgtgtgct acaatgagcc caaggtaaca 120
acaagctgcc cccagcaggg tctgcaggct gtgcccactg gcatcccagc ctctagccag 180
cgaatcttcc tgcatggcaa ccgaatctct cacgtgccag ctgcgagctt ccagtcatgc 240
cgaaatctca ctatcctgtg gctgcactct aatgcgctgg ctcggatcga tgctgctgcc 300
ttcactggtc tgaccctcct ggagcaacta gatcttagtg ataatgcaca gcttcatgtc 360
gtggacccta ccacgttcca cggcctgggc cacctgcaca cactgcacct agaccgatgt 420
ggcctgcggg agctgggtcc cggcctattc cgtggactag cagctctgca gtacctctac 480
ctacaagaca acaatctgca ggcactccct gacaacacct ttcgagacct gggcaacctc 540
acgcatctct ttctgcatgg caaccgtatc cccagtgtgc ctgagcacgc tttccgtggc 600
ctgcacagtc ttgaccgcct cctcttgcac cagaaccatg tggctcgtgt gcacccacat 660
gccttccggg accttggccg cctcatgacc ctctacctgt ttgccaacaa cctctccatg 720
ctgcctgcag aggtcctaat gcccctgagg tctctgcagt acctgcgact caatgacaac 780
ccctgggtgt gtgactgccg ggcacgtcca ctctgggcct ggctgcagaa gttccgaggt 840
tcctcatcag aggtgccctg caacctgccc caacgcctgg cagaccgtga tcttaagcgc 900
ctcgctgcca gtgacctaga gggctgtgct gtggcttcag gacccttccg tcccatccag 960
accagtcagc tcactgatga ggagctgctg agcctcccca agtgctgcca gccagatgct 1020
gcagacaaag cctcagtact ggaacccggg aggccagctt ctgccggaaa cgccctcaag 1080
ggacgtgtgc ctcccggtga cactccacca ggcaatggct caggccctcg gcacatcaat 1140
gactctccat ttggaacttt gcccagctct gcagagcccc cactgactgc cctgcggcct 1200
gggggttccg agccaccagg acttcccacc actggtcccc gcaggaggcc aggttgttcc 1260
cggaagaatc gcacccgcag ccactgccgt ctgggccagg cgggaagtgg ggccagtgga 1320
acaggggacg cagagggttc aggggctctg cctgctctgg cctgcagcct tgctcctctg 1380
ggccttgcac tggtactttg gacagtgctt gggccctgct ga 1422

<210> 203
<211> 473
<212> PRT
<213> Mus musculus

<400> 203
Met Lys Arg Ala Ser Ser Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1 5 10 15

Trp Leu Gln Val Trp Arg Val Ala Thr Pro Cys Pro Gly Ala Cys Val
20 25 30

Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser Cys Pro Gln Gln Gly Leu
35 40 45

Gln Ala Val Pro Thr Gly Ile Pro Ala Ser Ser Gln Arg Ile Phe Leu
50 55 60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Gln Ser Cys
65 70 75 80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Ala Leu Ala Arg Ile
85 90 95

Asp Ala Ala Ala Phe Thr Gly Leu Thr Leu Leu Glu Gln Leu Asp Leu
100 105 110

Ser Asp Asn Ala Gln Leu His Val Val Asp Pro Thr Thr Phe His Gly
115 120 125

Leu Gly His Leu His Thr Leu His Leu Asp Arg Cys Gly Leu Arg Glu
130 135 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145 150 155 160

Leu Gln Asp Asn Asn Leu Gln Ala Leu Pro Asp Asn Thr Phe Arg Asp
165 170 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Pro Ser
180 185 190

Val Pro Glu His Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
195 200 205

Leu His Gln Asn His Val Ala Arg Val His Pro His Ala Phe Arg Asp
210 215 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Met
225 230 235 240

Leu Pro Ala Glu Val Leu Met Pro Leu Arg Ser Leu Gln Tyr Leu Arg
245 250 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
260 265 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Asn
275 280 285

Leu Pro Gln Arg Leu Ala Asp Arg Asp Leu Lys Arg Leu Ala Ala Ser
290 295 300

Asp Leu Glu Gly Cys Ala Val Ala Ser Gly Pro Phe Arg Pro Ile Gln
305 310 315 320

Thr Ser Gln Leu Thr Asp Glu Glu Leu Leu Ser Leu Pro Lys Cys Cys

325 330 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
340 345 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Thr
355 360 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
370 375 380

Gly Thr Leu Pro Ser Ser Ala Glu Pro Pro Leu Thr Ala Leu Arg Pro
385 390 395 400

Gly Gly Ser Glu Pro Pro Gly Leu Pro Thr Thr Gly Pro Arg Arg Arg
405 410 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
420 425 430

Gln Ala Gly Ser Gly Ala Ser Gly Thr Gly Asp Ala Glu Gly Ser Gly
435 440 445

Ala Leu Pro Ala Leu Ala Cys Ser Leu Ala Pro Leu Gly Leu Ala Leu
450 455 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465 470

<210> 204
<211> 1422
<212> DNA
<213> Mus musculus

<400> 204
atgaagaggg cgtcctccgg aggaagcagg ctgctggcat gggtgttatg gctacaggcc 60
tggagggtag caacaccatg ccctggtgct tgtgtgtgct acaatgagcc caaggtatca 120
acaagctgcc cccagcaggg tctgcaggct gtgcccactg gcatcccagc ctctagccag 180
cgaatcttcc tgcatggcaa ccgaatctct cacgtgccag ctgcgagctt ccagtcatgc 240
cgaaatctca ctatcctgtg gctgcactct aatgcgctgg ctcggatcga tgctgctgcc 300
ttcactggtc tgaccctcct ggagcaacta gatcttagtg ataatgcaca gcttcatgtc 360
gtggacccta ccacgttcca cggcctgggc cacctgcaca cactgcacct agaccgatgt 420
ggcctgcggg agctgggtcc cggcctattc cgtggactag cagctctgca gtacctctac 480
ctacaagaca acaatctgca ggcactccct gacaacacct ttcgagacct gggcaacctc 540
acgcatctct ttctgcatgg caaccgtatc cccagtgtgc ctgagcacgc tttccgtggc 600
ctgcacagtc ttgaccgcct cctcttgcac cagaaccatg tggctcgtgt gcacccacat 660
gccttccggg accttggccg cctcatgacc ctctacctgt ttgccaacaa cctctccatg 720
ctgcctgcag aggtcctaat gcccctgagg tctctgcagt acctgcgact caatgacaac 780
ccctgggtgt gtgactgccg ggcacgtcca ctctgggcct ggctgcagaa gttccgaggt 840
tcctcatcag aggtgccctg caacctgccc caacgcctgg cagaccgtga tcttaagcgc 900
ctcgctgcca gtgacctaga gggctgtgct gtggcttcag gacccttccg tcccatccag 960
accagtcagc tcactgatga ggagctgctg agcctcccca agtgctgcca gccagatgct 1020
gcagacaaag cctcagtact ggaacccggg aggccagctt ctgccggaaa cgccctcaag 1080
ggacgtgtgc ctcccggtga cactccacca ggcaatggct caggccctcg gcacatcaat 1140
gactctccat ttggaacttt gcccagctct gcagagcccc cactgactgc cctgcggcct 1200
gggggttccg agccaccagg acttcccacc actggtcccc gcaggaggcc aggttgttcc 1260
cggaagaatc gcacccgcag ccactgccgt ctgggccagg cgggaagtgg ggccagtgga 1320

acaggggacg cagagggttc aggggctctg cctgctctgg cctgcagcct tgctcctctg 1380
ggccttgcac tggtactttg gacagtgctt gggccctgct ga 1422

<210> 205
<211> 473
<212> PRT
<213> Mus musculus

<400> 205
Met Lys Arg Ala Ser Ser Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1 5 10 15

Trp Leu Gln Ala Trp Arg Val Ala Thr Pro Cys Pro Gly Ala Cys Val
20 25 30

Cys Tyr Asn Glu Pro Lys Val Ser Thr Ser Cys Pro Gln Gln Gly Leu
35 40 45

Gln Ala Val Pro Thr Gly Ile Pro Ala Ser Ser Gln Arg Ile Phe Leu
50 55 60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Gln Ser Cys
65 70 75 80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Ala Leu Ala Arg Ile
85 90 95

Asp Ala Ala Ala Phe Thr Gly Leu Thr Leu Leu Glu Gln Leu Asp Leu
100 105 110

Ser Asp Asn Ala Gln Leu His Val Val Asp Pro Thr Thr Phe His Gly
115 120 125

Leu Gly His Leu His Thr Leu His Leu Asp Arg Cys Gly Leu Arg Glu
130 135 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145 150 155 160

Leu Gln Asp Asn Asn Leu Gln Ala Leu Pro Asp Asn Thr Phe Arg Asp
165 170 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Pro Ser
180 185 190

Val Pro Glu His Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
195 200 205

Leu His Gln Asn His Val Ala Arg Val His Pro His Ala Phe Arg Asp
210 215 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Met
225 230 235 240

Leu Pro Ala Glu Val Leu Met Pro Leu Arg Ser Leu Gln Tyr Leu Arg
245 250 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp

260 265 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Asn
275 280 285

Leu Pro Gln Arg Leu Ala Asp Arg Asp Leu Lys Arg Leu Ala Ala Ser
290 295 300

Asp Leu Glu Gly Cys Ala Val Ala Ser Gly Pro Phe Arg Pro Ile Gln
305 310 315 320

Thr Ser Gln Leu Thr Asp Glu Glu Leu Leu Ser Leu Pro Lys Cys Cys
325 330 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
340 345 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Thr
355 360 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
370 375 380

Gly Thr Leu Pro Ser Ser Ala Glu Pro Pro Leu Thr Ala Leu Arg Pro
385 390 395 400

Gly Gly Ser Glu Pro Pro Gly Leu Pro Thr Thr Gly Pro Arg Arg Arg
405 410 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
420 425 430

Gln Ala Gly Ser Gly Ala Ser Gly Thr Gly Asp Ala Glu Gly Ser Gly
435 440 445

Ala Leu Pro Ala Leu Ala Cys Ser Leu Ala Pro Leu Gly Leu Ala Leu
450 455 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465 470

<210> 206
<211> 1422
<212> DNA
<213> Mus musculus

<400> 206
atgaagaggg cgtcctccgg aggaagcagg ctgctggcat gggtgttatg gctacaggcc 60
tggagggtag caacaccatg ccctggtgct tgtgtgtgct acaatgagcc caaggtaaca 120
acaagctgcc cccagcaggg tctgcaggct gtgcccactg gcatcccagc ctctagcgag 180
cgaatcttcc tgcatggcaa ccgaatctct cacgtgccag ctgcgagctt ccagtcatgc 240
cgaaatctca ctatcctgtg gctgcactct aatgcgctgg ctcggatcga tgctgctgcc 300
ttcactggtc tgaccctcct ggagcaacta gatcttagtg ataatgcaca gcttcatgtc 360
gtggacccta ccacgttcca cggcctgggc cacctgcaca cactgcacct agaccgatgt 420
ggcctgcggg agctgggtcc cggcctattc cgtggactag cagctctgca gtacctctac 480
ctacaagaca acaatctgca ggcactccct gacaacacct ttcgagacct gggcaacctc 540
acgcatctct ttctgcatgg caaccgtatc cccagtgtgc ctgagcacgc tttccgtggc 600

```
ctgcacagtc ttgaccgcct cctcttgcac cagaaccatg tggctcgtgt gcacccacat 660
gccttccggg accttggccg cctcatgacc ctctacctgt ttgccaacaa cctctccatg 720
ctgcctgcag aggtcctaat gcccctgagg tctctgcagt acctgcgact caatgacaac 780
ccctgggtgt gtgactgccg ggcacgtcca ctctgggcct ggctgcagaa gttccgaggt 840
tcctcatcag aggtgccctg caacctgccc caacgcctgg cagaccgtga tcttaagcgc 900
ctcgctgcca gtgacctaga gggctgtgct gtggcttcag gacccttccg tcccatccag 960
accagtcagc tcactgatga ggagctgctg agcctcccca agtgctgcca gccagatgct 1020
gcagacaaag cctcagtact ggaacccggg aggccagctt ctgccggaaa cgccctcaag 1080
ggacgtgtgc ctcccggtga cactccacca ggcaatggct caggccctcg gcacatcaat 1140
gactctccat ttggaacttt gcccagctct gcagagcccc cactgactgc cctgcggcct 1200
gggggttccg agccaccagg acttcccacc actggtcccc gcaggaggcc aggttgttcc 1260
cggaagaatc gcacccgcag ccactgccgt ctgggccagg cgggaagtgg ggccagtgga 1320
acaggggacg cagagggttc aggggctctg cctgctctgg cctgcagcct tgctcctctg 1380
ggccttgcac tggtactttg gacagtgctt gggccctgct ga                    1422
```

<210> 207
<211> 473
<212> PRT
<213> Mus musculus

<400> 207

```
Met Lys Arg Ala Ser Ser Gly Gly Ser Arg Leu Leu Ala Trp Val Leu
1               5                   10                  15

Trp Leu Gln Ala Trp Arg Val Ala Thr Pro Cys Pro Gly Ala Cys Val
            20                  25                  30

Cys Tyr Asn Glu Pro Lys Val Thr Thr Ser Cys Pro Gln Gln Gly Leu
        35                  40                  45

Gln Ala Val Pro Thr Gly Ile Pro Ala Ser Ser Glu Arg Ile Phe Leu
    50                  55                  60

His Gly Asn Arg Ile Ser His Val Pro Ala Ala Ser Phe Gln Ser Cys
65                  70                  75                  80

Arg Asn Leu Thr Ile Leu Trp Leu His Ser Asn Ala Leu Ala Arg Ile
                85                  90                  95

Asp Ala Ala Ala Phe Thr Gly Leu Thr Leu Leu Glu Gln Leu Asp Leu
            100                 105                 110

Ser Asp Asn Ala Gln Leu His Val Val Asp Pro Thr Thr Phe His Gly
        115                 120                 125

Leu Gly His Leu His Thr Leu His Leu Asp Arg Cys Gly Leu Arg Glu
    130                 135                 140

Leu Gly Pro Gly Leu Phe Arg Gly Leu Ala Ala Leu Gln Tyr Leu Tyr
145                 150                 155                 160

Leu Gln Asp Asn Asn Leu Gln Ala Leu Pro Asp Asn Thr Phe Arg Asp
                165                 170                 175

Leu Gly Asn Leu Thr His Leu Phe Leu His Gly Asn Arg Ile Pro Ser
            180                 185                 190

Val Pro Glu His Ala Phe Arg Gly Leu His Ser Leu Asp Arg Leu Leu
```

```
                195                 200                 205

Leu His Gln Asn His Val Ala Arg Val His Pro His Ala Phe Arg Asp
    210                 215                 220

Leu Gly Arg Leu Met Thr Leu Tyr Leu Phe Ala Asn Asn Leu Ser Met
225                 230                 235                 240

Leu Pro Ala Glu Val Leu Met Pro Leu Arg Ser Leu Gln Tyr Leu Arg
                245                 250                 255

Leu Asn Asp Asn Pro Trp Val Cys Asp Cys Arg Ala Arg Pro Leu Trp
            260                 265                 270

Ala Trp Leu Gln Lys Phe Arg Gly Ser Ser Ser Glu Val Pro Cys Asn
        275                 280                 285

Leu Pro Gln Arg Leu Ala Asp Arg Asp Leu Lys Arg Leu Ala Ala Ser
    290                 295                 300

Asp Leu Glu Gly Cys Ala Val Ala Ser Gly Pro Phe Arg Pro Ile Gln
305                 310                 315                 320

Thr Ser Gln Leu Thr Asp Glu Glu Leu Leu Ser Leu Pro Lys Cys Cys
                325                 330                 335

Gln Pro Asp Ala Ala Asp Lys Ala Ser Val Leu Glu Pro Gly Arg Pro
            340                 345                 350

Ala Ser Ala Gly Asn Ala Leu Lys Gly Arg Val Pro Pro Gly Asp Thr
        355                 360                 365

Pro Pro Gly Asn Gly Ser Gly Pro Arg His Ile Asn Asp Ser Pro Phe
    370                 375                 380

Gly Thr Leu Pro Ser Ser Ala Glu Pro Pro Leu Thr Ala Leu Arg Pro
385                 390                 395                 400

Gly Gly Ser Glu Pro Pro Gly Leu Pro Thr Thr Gly Pro Arg Arg Arg
                405                 410                 415

Pro Gly Cys Ser Arg Lys Asn Arg Thr Arg Ser His Cys Arg Leu Gly
            420                 425                 430

Gln Ala Gly Ser Gly Ala Ser Gly Thr Gly Asp Ala Glu Gly Ser Gly
        435                 440                 445

Ala Leu Pro Ala Leu Ala Cys Ser Leu Ala Pro Leu Gly Leu Ala Leu
    450                 455                 460

Val Leu Trp Thr Val Leu Gly Pro Cys
465                 470


<210> 208
<211> 624
<212> DNA
<213> Homo sapiens
```

<400> 208
atggagaatg aagatgggta tatgacggtg agtttcaaga atcgttgtaa atcgaagcag 60
aaatctaaag atttctccct atatccacaa tattattgtc ttctgctcat atttggatgc 120
attgtgatcc ttatattcat tatgacaggg attgacctga agttctggca taaaaaaatg 180
gatttctccc agaatgtaaa catcagcagt ctatcaggac acaattactt gtgcccaaat 240
gactggctgt tgaacgaagg gaaatgttac tggttttcaa cttcttttaa aacgtggaaa 300
gagagtcaac gtgattgtac acagctacag gcacatttac tggtgattca aaatttggat 360
gagctggagt tcatacagaa cagtttaaaa cctggacatt ttggttggat tggactatat 420
gttacattcc aagggaacct atggatgtgg atagatgaac actttttagt tccagaattg 480
ttttcagtga ttggaccaac tgatgacagg agctgtgccg ttatcacagg aaactgggtg 540
tattctgaag actgtagctc cac[illegible]tttaag ggcatttgcc agagagatgc gatcttgacg 600
cacaatggaa ccagtggtgt gtaa 624

<210> 209
<211> 207
<212> PRT
<213> Homo sapiens

<400> 209
Met Glu Asn Glu Asp Gly Tyr Met Thr Val Ser Phe Lys Asn Arg Cys
1 5 10 15

Lys Ser Lys Gln Lys Ser Lys Asp Phe Ser Leu Tyr Pro Gln Tyr Tyr
20 25 30

Cys Leu Leu Leu Ile Phe Gly Cys Ile Val Ile Leu Ile Phe Ile Met
35 40 45

Thr Gly Ile Asp Leu Lys Phe Trp His Lys Lys Met Asp Phe Ser Gln
50 55 60

Asn Val Asn Ile Ser Ser Leu Ser Gly His Asn Tyr Leu Cys Pro Asn
65 70 75 80

Asp Trp Leu Leu Asn Glu Gly Lys Cys Tyr Trp Phe Ser Thr Ser Phe
85 90 95

Lys Thr Trp Lys Glu Ser Gln Arg Asp Cys Thr Gln Leu Gln Ala His
100 105 110

Leu Leu Val Ile Gln Asn Leu Asp Glu Leu Glu Phe Ile Gln Asn Ser
115 120 125

Leu Lys Pro Gly His Phe Gly Trp Ile Gly Leu Tyr Val Thr Phe Gln
130 135 140

Gly Asn Leu Trp Met Trp Ile Asp Glu His Phe Leu Val Pro Glu Leu
145 150 155 160

Phe Ser Val Ile Gly Pro Thr Asp Asp Arg Ser Cys Ala Val Ile Thr
165 170 175

Gly Asn Trp Val Tyr Ser Glu Asp Cys Ser Ser Thr Phe Lys Gly Ile
180 185 190

Cys Gln Arg Asp Ala Ile Leu Thr His Asn Gly Thr Ser Gly Val
195 200 205

<210> 210
<211> 624
<212> DNA
<213> Homo sapiens

<400> 210
```
atggagaatg aagatgggta tatgacgctg agtttcaaga atcgttgtaa atcgaaggag 60
aaatctaaag atttctccct atatccacaa tattattgtc ttctgctcat atttggatgc 120
attgtgatcc ttatattcat tatgacaggg attgacctga agttctggca taaaaaaatg 180
gatttctccc agaatgtaaa catcagcagt ctatcaggac acaattactt gtgcccaaat 240
gactggctgt tgaacgaagg gaaatgttac tggttttcaa cttcttttaa aacgtggaaa 300
gagagtcaac gtgattgtac acagctacag gcacatttac tggtgattca aaatttggat 360
gagctggagt tcatacagaa cagtttaaaa cctggacatt ttggttggat tggactatat 420
gttacattcc aagggaacct atggatgtgg atagatgaac actttttagt tccagaattg 480
ttttcagtga ttggaccaac tgatgacagg agctgtgccg ttatcacagg aaactgggtg 540
tattctgaag actgtagctc cacatttaag ggcatttgcc agagagatgc gatcttgacg 600
cacaatggaa ccagtggtgt gtaa                                        624
```

<210> 211
<211> 207
<212> PRT
<213> Homo sapiens

<400> 211
```
Met Glu Asn Glu Asp Gly Tyr Met Thr Leu Ser Phe Lys Asn Arg Cys
1               5                   10                  15

Lys Ser Lys Glu Lys Ser Lys Asp Phe Ser Leu Tyr Pro Gln Tyr Tyr
            20                  25                  30

Cys Leu Leu Leu Ile Phe Gly Cys Ile Val Ile Leu Ile Phe Ile Met
        35                  40                  45

Thr Gly Ile Asp Leu Lys Phe Trp His Lys Lys Met Asp Phe Ser Gln
    50                  55                  60

Asn Val Asn Ile Ser Ser Leu Ser Gly His Asn Tyr Leu Cys Pro Asn
65                  70                  75                  80

Asp Trp Leu Leu Asn Glu Gly Lys Cys Tyr Trp Phe Ser Thr Ser Phe
                85                  90                  95

Lys Thr Trp Lys Glu Ser Gln Arg Asp Cys Thr Gln Leu Gln Ala His
            100                 105                 110

Leu Leu Val Ile Gln Asn Leu Asp Glu Leu Glu Phe Ile Gln Asn Ser
        115                 120                 125

Leu Lys Pro Gly His Phe Gly Trp Ile Gly Leu Tyr Val Thr Phe Gln
    130                 135                 140

Gly Asn Leu Trp Met Trp Ile Asp Glu His Phe Leu Val Pro Glu Leu
145                 150                 155                 160

Phe Ser Val Ile Gly Pro Thr Asp Asp Arg Ser Cys Ala Val Ile Thr
                165                 170                 175

Gly Asn Trp Val Tyr Ser Glu Asp Cys Ser Ser Thr Phe Lys Gly Ile
```

```
            180                 185                 190

Cys Gln Arg Asp Ala Ile Leu Thr His Asn Gly Thr Ser Gly Val
        195                 200                 205


<210> 212
<211> 624
<212> DNA
<213> Homo sapiens

<400> 212
atggagaatg aagatgggta tatgacgctg agtttcaaga atcgttgtaa atcgaagcag 60
agatctaaag atttctccct atatccacaa tattattgtc ttctgctcat atttggatgc 120
attgtgatcc ttatattcat tatgacaggg attgacctga agttctggca taaaaaaatg 180
gatttctccc agaatgtaaa catcagcagt ctatcaggac acaattactt gtgcccaaat 240
gactggctgt tgaacgaagg gaaatgttac tggttttcaa cttcttttaa aacgtggaaa 300
gagagtcaac gtgattgtac acagctacag gcacatttac tggtgattca aaatttggat 360
gagctggagt tcatacagaa cagtttaaaa cctggacatt ttggttggat tggactatat 420
gttacattcc aagggaacct atggatgtgg atagatgaac actttttagt tccagaattg 480
ttttcagtga ttggaccaac tgatgacagg agctgtgccg ttatcacagg aaactgggtg 540
tattctgaag actgtagctc cacatttaag ggcatttgcc agagagatgc gatcttgacg 600
cacaatggaa ccagtggtgt gtaa                                        624

<210> 213
<211> 207
<212> PRT
<213> Homo sapiens

<400> 213
Met Glu Asn Glu Asp Gly Tyr Met Thr Leu Ser Phe Lys Asn Arg Cys
1               5                   10                  15

Lys Ser Lys Gln Arg Ser Lys Asp Phe Ser Leu Tyr Pro Gln Tyr Tyr
            20                  25                  30

Cys Leu Leu Leu Ile Phe Gly Cys Ile Val Ile Leu Ile Phe Ile Met
        35                  40                  45

Thr Gly Ile Asp Leu Lys Phe Trp His Lys Lys Met Asp Phe Ser Gln
    50                  55                  60

Asn Val Asn Ile Ser Ser Leu Ser Gly His Asn Tyr Leu Cys Pro Asn
65                  70                  75                  80

Asp Trp Leu Leu Asn Glu Gly Lys Cys Tyr Trp Phe Ser Thr Ser Phe
                85                  90                  95

Lys Thr Trp Lys Glu Ser Gln Arg Asp Cys Thr Gln Leu Gln Ala His
            100                 105                 110

Leu Leu Val Ile Gln Asn Leu Asp Glu Leu Glu Phe Ile Gln Asn Ser
        115                 120                 125

Leu Lys Pro Gly His Phe Gly Trp Ile Gly Leu Tyr Val Thr Phe Gln
    130                 135                 140

Gly Asn Leu Trp Met Trp Ile Asp Glu His Phe Leu Val Pro Glu Leu
```

145 150 155 160

Phe Ser Val Ile Gly Pro Thr Asp Asp Arg Ser Cys Ala Val Ile Thr
165 170 175

Gly Asn Trp Val Tyr Ser Glu Asp Cys Ser Ser Thr Phe Lys Gly Ile
180 185 190

Cys Gln Arg Asp Ala Ile Leu Thr His Asn Gly Thr Ser Gly Val
195 200 205

<210> 214
<211> 624
<212> DNA
<213> Homo sapiens

<400> 214
atggagaatg aagatgggta tatgacgctg agtttcaaga atcgttgtaa atcgaagcag 60
aaaactaaag atttctccct atatccacaa tattattgtc ttctgctcat atttggatgc 120
attgtgatcc ttatattcat tatgacaggg attgacctga agttctggca taaaaaaatg 180
gatttctccc agaatgtaaa catcagcagt ctatcaggac acaattactt gtgcccaaat 240
gactggctgt tgaacgaagg gaaatgttac tggttttcaa cttcttttaa aacgtggaaa 300
gagagtcaac gtgattgtac acagctacag gcacatttac tggtgattca aaatttggat 360
gagctggagt tcatacagaa cagtttaaaa cctggacatt ttggttggat tggactatat 420
gttacattcc aagggaacct atggatgtgg atagatgaac actttttagt tccagaattg 480
ttttcagtga ttggaccaac tgatgacagg agctgtgccg ttatcacagg aaactgggtg 540
tattctgaag actgtagctc cacatttaag ggcatttgcc agagagatgc gatcttgacg 600
cacaatggaa ccagtggtgt gtaa 624

<210> 215
<211> 207
<212> PRT
<213> Homo sapiens

<400> 215
Met Glu Asn Glu Asp Gly Tyr Met Thr Leu Ser Phe Lys Asn Arg Cys
1 5 10 15

Lys Ser Lys Gln Lys Thr Lys Asp Phe Ser Leu Tyr Pro Gln Tyr Tyr
20 25 30

Cys Leu Leu Leu Ile Phe Gly Cys Ile Val Ile Leu Ile Phe Ile Met
35 40 45

Thr Gly Ile Asp Leu Lys Phe Trp His Lys Lys Met Asp Phe Ser Gln
50 55 60

Asn Val Asn Ile Ser Ser Leu Ser Gly His Asn Tyr Leu Cys Pro Asn
65 70 75 80

Asp Trp Leu Leu Asn Glu Gly Lys Cys Tyr Trp Phe Ser Thr Ser Phe
85 90 95

Lys Thr Trp Lys Glu Ser Gln Arg Asp Cys Thr Gln Leu Gln Ala His
100 105 110

Leu Leu Val Ile Gln Asn Leu Asp Glu Leu Glu Phe Ile Gln Asn Ser

115 120 125

Leu Lys Pro Gly His Phe Gly Trp Ile Gly Leu Tyr Val Thr Phe Gln
130 135 140

Gly Asn Leu Trp Met Trp Ile Asp Glu His Phe Leu Val Pro Glu Leu
145 150 155 160

Phe Ser Val Ile Gly Pro Thr Asp Asp Arg Ser Cys Ala Val Ile Thr
165 170 175

Gly Asn Trp Val Tyr Ser Glu Asp Cys Ser Ser Thr Phe Lys Gly Ile
180 185 190

Cys Gln Arg Asp Ala Ile Leu Thr His Asn Gly Thr Ser Gly Val
195 200 205

<210> 216
<211> 183
<212> DNA
<213> Homo sapiens

<400> 216
atgtattctt ttatctgtat cctgcctctc ttgctcttgg cttcctgcct tctctcctac 60
tcatttttag aacagtctag atgcaggcag ctagaagagt tatttcctcc aagctgtcta 120
ggaaaaggga caattaaaga gagattctgc acttattatg atataaaaaa agaaaaacaa 180
tga 183

<210> 217
<211> 60
<212> PRT
<213> Homo sapiens

<400> 217
Met Tyr Ser Phe Ile Cys Ile Leu Pro Leu Leu Leu Leu Ala Ser Cys
1 5 10 15

Leu Leu Ser Tyr Ser Phe Leu Glu Gln Ser Arg Cys Arg Gln Leu Glu
20 25 30

Glu Leu Phe Pro Pro Ser Cys Leu Gly Lys Gly Thr Ile Lys Glu Arg
35 40 45

Phe Cys Thr Tyr Tyr Asp Ile Lys Lys Glu Lys Gln
50 55 60

<210> 218
<211> 183
<212> DNA
<213> Homo sapiens

<400> 218
atgtattctt ttctctgtat cctgcctctc ttgctcttgg cttcctgcct tctctccttc 60
tcatttttag aacagtctag atgcaggcag ctagaagagt tatttcctcc aagctgtcta 120
ggaaaaggga caattaaaga gagattctgc acttattatg atataaaaaa agaaaaacaa 180
tga 183

<210> 219
<211> 60
<212> PRT
<213> Homo sapiens

<400> 219
Met Tyr Ser Phe Leu Cys Ile Leu Pro Leu Leu Leu Leu Ala Ser Cys
1 5 10 15

Leu Leu Ser Phe Ser Phe Leu Glu Gln Ser Arg Cys Arg Gln Leu Glu
20 25 30

Glu Leu Phe Pro Pro Ser Cys Leu Gly Lys Gly Thr Ile Lys Glu Arg
35 40 45

Phe Cys Thr Tyr Tyr Asp Ile Lys Lys Glu Lys Gln
50 55 60

<210> 220
<211> 183
<212> DNA
<213> Homo sapiens

<400> 220
atgtattctt ttctctgtat cctgcctctc ttgctcttgg cttcctgcct tctctcctac 60
acatttttag aacagtctag atgcaggcag ctagaagagt tatttcctcc aagctgtcta 120
ggaaaaggga caattaaaga gagattctgc acttattatg atataaaaaa agaaaaacaa 180
tga 183

<210> 221
<211> 60
<212> PRT
<213> Homo sapiens

<400> 221
Met Tyr Ser Phe Leu Cys Ile Leu Pro Leu Leu Leu Leu Ala Ser Cys
1 5 10 15

Leu Leu Ser Tyr Thr Phe Leu Glu Gln Ser Arg Cys Arg Gln Leu Glu
20 25 30

Glu Leu Phe Pro Pro Ser Cys Leu Gly Lys Gly Thr Ile Lys Glu Arg
35 40 45

Phe Cys Thr Tyr Tyr Asp Ile Lys Lys Glu Lys Gln
50 55 60

<210> 222
<211> 183
<212> DNA
<213> Homo sapiens

<400> 222
atgtattctt ttctctgtat cctgcctctc ttgctcttgg cttcctgcct tctctcctac 60
tcatttttag aacagtctaa atgcaggcag ctagaagagt tatttcctcc aagctgtcta 120
ggaaaaggga caattaaaga gagattctgc acttattatg atataaaaaa agaaaaacaa 180

tga 183

<210> 223
<211> 60
<212> PRT
<213> Homo sapiens

<400> 223
Met Tyr Ser Phe Leu Cys Ile Leu Pro Leu Leu Leu Leu Ala Ser Cys
1 5 10 15

Leu Leu Ser Tyr Ser Phe Leu Glu Gln Ser Lys Cys Arg Gln Leu Glu
20 25 30

Glu Leu Phe Pro Pro Ser Cys Leu Gly Lys Gly Thr Ile Lys Glu Arg
35 40 45

Phe Cys Thr Tyr Tyr Asp Ile Lys Lys Glu Lys Gln
50 55 60

<210> 224
<211> 504
<212> DNA
<213> Homo sapiens

<400> 224
atgtcagcgg gaacggttgt gatcactggc ggaatcctag ctacggtgat cctcctctgc 60
atcattgccg tcctgtgcta ctgcaggctc cagtattact gctgcaagaa gagcggaacc 120
gaggttgcag acgaggagga ggagcgggag cacgaccttc ccacgcatcc cagaggcccc 180
acctgcaatg cctgcagctc ccaagccctg gacggcagag gcagcctggc gcctctcacc 240
agcgagccct gcagccagcc ctgtggggtg gccgcgagcc actgcactac ctgctcccca 300
tacagctccc cctttacat acggacggct gacatggtgc ccaatggggg tggaggcgag 360
aggctctcct ttgctcccac atactacaaa gagggggac ccccatccct caaattggca 420
gcaccccaga gttacccggt gacctggcca ggctctgggc gtgaggcctt caccaatcca 480
agggctatta gtacagacgt gtaa 504

<210> 225
<211> 167
<212> PRT
<213> Homo sapiens

<400> 225
Met Ser Ala Gly Thr Val Val Ile Thr Gly Gly Ile Leu Ala Thr Val
1 5 10 15

Ile Leu Leu Cys Ile Ile Ala Val Leu Cys Tyr Cys Arg Leu Gln Tyr
20 25 30

Tyr Cys Cys Lys Lys Ser Gly Thr Glu Val Ala Asp Glu Glu Glu Glu
35 40 45

Arg Glu His Asp Leu Pro Thr His Pro Arg Gly Pro Thr Cys Asn Ala
50 55 60

Cys Ser Ser Gln Ala Leu Asp Gly Arg Gly Ser Leu Ala Pro Leu Thr
65 70 75 80

Ser Glu Pro Cys Ser Gln Pro Cys Gly Val Ala Ala Ser His Cys Thr
85 90 95

Thr Cys Ser Pro Tyr Ser Ser Pro Phe Tyr Ile Arg Thr Ala Asp Met
100 105 110

Val Pro Asn Gly Gly Gly Gly Glu Arg Leu Ser Phe Ala Pro Thr Tyr
115 120 125

Tyr Lys Glu Gly Gly Pro Pro Ser Leu Lys Leu Ala Ala Pro Gln Ser
130 135 140

Tyr Pro Val Thr Trp Pro Gly Ser Gly Arg Glu Ala Phe Thr Asn Pro
145 150 155 160

Arg Ala Ile Ser Thr Asp Val
165

<210> 226
<211> 504
<212> DNA
<213> Homo sapiens

<400> 226
atgacagcgg gaacggttgt gatcactggc ggaatcctag ctacggtgat cctcctctgc 60
ctcattgccg tcctgtgcta ctgcaggctc cagtattact gctgcaagaa gagcggaacc 120
gaggttgcag acgaggagga ggagcgggag cacgaccttc ccacgcatcc cagaggcccc 180
acctgcaatg cctgcagctc ccaagccctg gacggcagag gcagcctggc gcctctcacc 240
agcgagccct gcagccagcc ctgtggggtg gccgcgagcc actgcactac ctgctcccca 300
tacagctccc ccttttacat acggacggct gacatggtgc ccaatggggg tggaggcgag 360
aggctctcct ttgctcccac atactacaaa gagggggac ccccatccct caaattggca 420
gcaccccaga gttacccggt gacctggcca ggctctgggc gtgaggcctt caccaatcca 480
agggctatta gtacagacgt gtaa 504

<210> 227
<211> 167
<212> PRT
<213> Homo sapiens

<400> 227
Met Thr Ala Gly Thr Val Val Ile Thr Gly Gly Ile Leu Ala Thr Val
1 5 10 15

Ile Leu Leu Cys Leu Ile Ala Val Leu Cys Tyr Cys Arg Leu Gln Tyr
20 25 30

Tyr Cys Cys Lys Lys Ser Gly Thr Glu Val Ala Asp Glu Glu Glu Glu
35 40 45

Arg Glu His Asp Leu Pro Thr His Pro Arg Gly Pro Thr Cys Asn Ala
50 55 60

Cys Ser Ser Gln Ala Leu Asp Gly Arg Gly Ser Leu Ala Pro Leu Thr
65 70 75 80

Ser Glu Pro Cys Ser Gln Pro Cys Gly Val Ala Ala Ser His Cys Thr
85 90 95

Thr Cys Ser Pro Tyr Ser Ser Pro Phe Tyr Ile Arg Thr Ala Asp Met
100 105 110

Val Pro Asn Gly Gly Gly Gly Glu Arg Leu Ser Phe Ala Pro Thr Tyr
115 120 125

Tyr Lys Glu Gly Gly Pro Pro Ser Leu Lys Leu Ala Ala Pro Gln Ser
130 135 140

Tyr Pro Val Thr Trp Pro Gly Ser Gly Arg Glu Ala Phe Thr Asn Pro
145 150 155 160

Arg Ala Ile Ser Thr Asp Val
165

<210> 228
<211> 504
<212> DNA
<213> Homo sapiens

<400> 228
atgacagcgg gaacggttgt gatcactggc ggaatcctag ctacggtgat cctcctctgc 60
atcattgccg tcctgtgcta ctgcaagctc cagtattact gctgcaagaa gagcggaacc 120
gaggttgcag acgaggagga ggagcgggag cacgaccttc ccacgcatcc cagaggcccc 180
acctgcaatg cctgcagctc ccaagccctg gacggcagag gcagcctggc gcctctcacc 240
agcgagccct gcagccagcc ctgtggggtg gccgcgagcc actgcactac ctgctcccca 300
tacagctccc cctttacat acggacggct gacatggtgc ccaatggggg tggaggcgag 360
aggctctcct ttgctcccac atactacaaa gaggggggac ccccatccct caaattggca 420
gcaccccaga gttacccggt gacctggcca ggctctgggc gtgaggcctt caccaatcca 480
agggctatta gtacagacgt gtaa 504

<210> 229
<211> 167
<212> PRT
<213> Homo sapiens

<400> 229
Met Thr Ala Gly Thr Val Val Ile Thr Gly Gly Ile Leu Ala Thr Val
1 5 10 15

Ile Leu Leu Cys Ile Ile Ala Val Leu Cys Tyr Cys Lys Leu Gln Tyr
20 25 30

Tyr Cys Cys Lys Lys Ser Gly Thr Glu Val Ala Asp Glu Glu Glu Glu
35 40 45

Arg Glu His Asp Leu Pro Thr His Pro Arg Gly Pro Thr Cys Asn Ala
50 55 60

Cys Ser Ser Gln Ala Leu Asp Gly Arg Gly Ser Leu Ala Pro Leu Thr
65 70 75 80

Ser Glu Pro Cys Ser Gln Pro Cys Gly Val Ala Ala Ser His Cys Thr
85 90 95

Thr Cys Ser Pro Tyr Ser Ser Pro Phe Tyr Ile Arg Thr Ala Asp Met
100 105 110

Val Pro Asn Gly Gly Gly Gly Glu Arg Leu Ser Phe Ala Pro Thr Tyr
115 120 125

Tyr Lys Glu Gly Gly Pro Pro Ser Leu Lys Leu Ala Ala Pro Gln Ser
130 135 140

Tyr Pro Val Thr Trp Pro Gly Ser Gly Arg Glu Ala Phe Thr Asn Pro
145 150 155 160

Arg Ala Ile Ser Thr Asp Val
165

<210> 230
<211> 504
<212> DNA
<213> Homo sapiens

<400> 230
atgacagcgg gaacggttgt gatcactggc ggaatcctag ctacggtgat cctcctctgc 60
atcattgccg tcctgtgcta ctgcaggctc cagtattact gctgcaagaa gagcggaacc 120
gacgttgcag acgaggagga ggagcgggag cacgaccttc ccacgcatcc cagaggcccc 180
acctgcaatg cctgcagctc ccaagccctg gacggcagag gcagcctggc gcctctcacc 240
agcgagccct gcagccagcc ctgtggggtg gccgcgagcc actgcactac ctgctcccca 300
tacagctccc ccttttacat acggacggct gacatggtgc ccaatggggg tggaggcgag 360
aggctctcct ttgctcccac atactacaaa gagggggac ccccatccct caaattggca 420
gcacccccaga gttacccggt gacctggcca ggctctgggc gtgaggcctt caccaatcca 480
agggctatta gtacagacgt gtaa 504

<210> 231
<211> 167
<212> PRT
<213> Homo sapiens

<400> 231
Met Thr Ala Gly Thr Val Val Ile Thr Gly Gly Ile Leu Ala Thr Val
1 5 10 15

Ile Leu Leu Cys Ile Ile Ala Val Leu Cys Tyr Cys Arg Leu Gln Tyr
20 25 30

Tyr Cys Cys Lys Lys Ser Gly Thr Asp Val Ala Asp Glu Glu Glu Glu
35 40 45

Arg Glu His Asp Leu Pro Thr His Pro Arg Gly Pro Thr Cys Asn Ala
50 55 60

Cys Ser Ser Gln Ala Leu Asp Gly Arg Gly Ser Leu Ala Pro Leu Thr
65 70 75 80

Ser Glu Pro Cys Ser Gln Pro Cys Gly Val Ala Ala Ser His Cys Thr
85 90 95

Thr Cys Ser Pro Tyr Ser Ser Pro Phe Tyr Ile Arg Thr Ala Asp Met
100 105 110

Val Pro Asn Gly Gly Gly Gly Glu Arg Leu Ser Phe Ala Pro Thr Tyr
115 120 125

Tyr Lys Glu Gly Gly Pro Pro Ser Leu Lys Leu Ala Ala Pro Gln Ser
130 135 140

Tyr Pro Val Thr Trp Pro Gly Ser Gly Arg Glu Ala Phe Thr Asn Pro
145 150 155 160

Arg Ala Ile Ser Thr Asp Val
165

<210> 232
<211> 15
<212> PRT
<213> Homo sapiens

<400> 232
Cys Gly Asn Val Gly Leu Arg Ala Val Pro Leu Asp Leu Ala Gln
1 5 10 15

<210> 233
<211> 71
<212> PRT
<213> Homo sapiens

<400> 233
Glu Phe Arg Thr Pro Pro Thr Gly Cys Gly Phe Leu Leu Ala Ala Thr
1 5 10 15

Cys Leu Arg Gly Leu Lys Ser Val Gln Gln Asn Arg Val Trp Leu Cys
20 25 30

His Pro Gly Cys Ile Gly Glu Ile Ser Ala Gln Tyr Ser Leu Arg Ile
35 40 45

Leu Gly Ser Ser Asp Ser Ser Ala Ser Ala Ser Gln Val Pro Cys Cys
50 55 60

Arg Arg Arg Gly Trp Thr Arg
65 70

<210> 234
<211> 7
<212> PRT
<213> Homo sapiens

<400> 234
Arg Arg His His Pro Leu Gln
1 5

<210> 235
<211> 29
<212> PRT
<213> Homo sapiens

<400> 235

```
Ser Gly Pro Gly Val Glu Leu Ala Ser Gly His Val Arg Gly Lys Arg
  1               5                  10                  15

Glu Ala Gly Leu Tyr Ser Lys Ala Glu Ile Pro Leu Arg
            20                  25
```